Socioeconomic status and consumption in an emerging economy

Wagner A. Kamakura a,⁎, Jose A. Mazzon b

a Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708, USA
b Faculdade de Economia e Administração, Universidade de São Paulo, Avenida Prof. Luciano Gualberto, 908 CEP 05508-010, São Paulo, Brazil

A R T I C L E   I N F O

Article history:
First received in 29, June 2011 and was under review for 3 months
Available online 29 September 2012

Keywords:
Social-class
Socioeconomic status
Household consumption
Middle class

A B S T R A C T

Despite the central role of social class or socioeconomic status on consumer behavior and the fact that this construct has been utilized in marketing research for more than seven decades, the marketing literature is surprisingly lacking in the conceptualization and measurement of this important construct. In this study, we address these issues and propose a flexible and robust theory-based framework for socioeconomic stratification, which we apply to identify socio-economic strata during a period (2003 and 2009) of substantial economic and social development in one emerging economy (Brazil). We then use this stratification to examine the relationship between socioeconomic status and consumption. Our socioeconomic stratification framework shows how the recent economic development observed in Brazil benefited the lower strata, leading to the emergence of the country’s “new middle class.” We also find that despite the high income concentration still prevalent in Brazil, consumption in many product categories is more evenly distributed; therefore, firms would be ill-advised to follow a premium market positioning strategy targeted mostly to the upper classes because this would leave a substantial portion of the market to the competition.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

This study has three main purposes. The first is to propose a modeling framework for social stratification based on the concepts of social class and permanent income that is also flexible enough to accommodate the typical socio-demographic data collected in market research studies. The second is to apply the proposed framework to historical data from the emerging economy of Brazil, which represents an excellent “laboratory” for the study of socioeconomic status and for illustrating our proposed framework because of the social programs implemented there since 2001, which have led to dramatic socioeconomic shifts in that economy. The third purpose is to relate household consumption to social stratification to understand how consumption priorities vary across social strata, and to separate differences that are due to budget effects from those that truly reflect differences in consumption priorities.

1.1. Social stratification over time and across societies

Since the dawn of mankind, the question of social stratification has been a constant presence in advanced societies. In antiquity, Athenian society was stratified into Eupatrids, Metics and Slaves; Sparta had the Spartans, the Periecos and Serfs (Pomeroy, Burstein, Donlan, & Roberts, 1999); and ancient Rome was stratified into Patricians, Commoners and Slaves (Shelton, 1997). More recently, societies have been stratified into classes that might differ slightly in nomenclature but generally translate into upper class, upper middle class, middle class, working class and lower class (Beehley, 2004; Gilbert, 2002; Thompson & Hickey, 2005). In a number of developing countries the concept of social stratification has been prevalent as a way to relate socioeconomic status to consumption and for marketers to implement differentiated strategies (Corrales, Barberena, & Schmeichel, 2006). Socioeconomic status is important for market segmentation in developing countries because societies with emerging economies tend to be more hierarchical and exhibit a greater separation among the social classes, giving class distinctions a greater role than in economically developed societies (Burgess & Steenkamp, 2006). Moreover, global corporations have historically overlooked substantial sections of emerging markets because they were not economically viable and because of the firms’ focus on the most affluent strata of society. The recent growth of the so-called “middle class” in emerging economies calls for more precise and valid measurement of social classes. One applied example of the concept of socioeconomic status of consumers in South America is the “Criterio Brasil” developed by the Brazilian Association of Research Companies (ABEP) based on the possession of eight types of durable goods, the employment of domestic help, and the level of education attained by the head of household. This measurement is widely used in ad hoc marketing research, tracking studies, media and consumer research (ABEP, 2011). Similar efforts can be found across Central and South America (Corrales et al., 2006) based on different combinations of...
occupation and education of the head of household, the dwelling type and ownership status, the presence of a domestic worker and ownership of various durable goods. Marketers in several European countries also utilize standard criteria for social stratification, but the common practice in Europe (Urquijo & Lobl, 2003; Ware & Dinning, 2003) is to define social strata on the basis of the level of education (e.g., Portugal, Italy) and occupation (e.g., Portugal, Italy, Russia, France, UK, Germany) of the head of the household, as well as on household income (Italy).

The examples above of social stratification across many emerging and developed economies highlight the importance of social class or socioeconomic status (SES) as a basis for market segmentation and as an important common basis for integrating information across multiple sources (e.g., media, marketing research, retail sales and household panels).

1.2. Social stratification in marketing and other disciplines

Since its inception, the field of marketing has used social class as a tool for market segmentation (Bauer, Cunningham, & Worton, 1965; Coleman, 1983; Sivadas, 1997) and as a foundation for establishing a marketing communication strategy (Bass, Pessemier, & Tigert, 1969; Beakman, 1967; Rich & Jain, 1968). More recently, Williams (2002) studied how gender and social class affect consumers' buying criteria for a broad range of product categories; Dahl and Moreau (2007) utilized social class in their study of consumers' creative experiences; Rucker and Galinsky (2008) related social class to consumers' desire to acquire certain products; Steenkamp and de Jong (2010) found, in a global study, that people who self-classified in higher social classes have more positive attitudes toward global brands; Steenkamp, de Jong, and Baumgartner (2010) found, in a study of social-desirability bias in cross-national surveys, that among all the socio-demographic covariates they considered, social class was the only one with no impact on socially desirable responses. Most importantly to marketers, because it defines the position of individuals within a stratified social system, SES determines the context under which consumption occurs. The concept of SES is becoming even more important to marketers because of the rise of the BRICS (Brazil, Russia, India, China, South Africa); from 1995 to 2010, BRICS countries more than doubled their share of global GDP to over 15%, with a projected growth of their combined GDP from US$9 trillion to US$128 trillion between 2010 and 2050 (O'Neill, Wilson, Urushothaman, & Stupnytska, 2005). One of the driving forces for this growth is the strengthening of international markets (Renard, 2009) with the growth of a "middle class" with resources and hunger for consumption (Fioratti, 2006) and the reduction of poverty levels in these countries (Chan, Gabel, Jenner, & Schindele, 2011). The emergence of the "middle class" is viewed as a driver for economic growth because this vaguely-defined socioeconomic cohort strives for a better life-style and seeks the adoption of goods and services associated with higher social status (Cui & Song, 2009; Murphy, Shleifer, & Vishny, 1989), driving growth in internal demand for goods and services (Senauer & Goetz, 2003).

Socioeconomic status has also been a frequent and important segmentation criterion in research on education (Filmer & Pritchett, 2001), health (Deressa, Ali, & Berhanie, 2007; Gwatkin, Rusten, Johnson, & Wagstaff, 2000; Marmot, Bosma, Hemingway, Brunner, & Stansfeld, 1997; and Thomas, 2007), psychology (Gray-Little & Haf Dahl, 2000; Jayakody, Danziger, & Kessler, 1998; Rogler, 1996) and economic development (Sumarto, Suryadarma, & Suryahadi, 2007). There are many reasons for the central role of the concept of socioeconomic status across multiple disciplines. Socioeconomic status determines an individual's opportunities and challenges in all realms of life and provides information about an individual's access to social and economic resources (Duncan, Daly, McDonough, & Williams, 2002).

Despite the fact that social class has been widely utilized in marketing research over the past seven decades, relatively limited attention has been paid to the construct itself. This dearth of research on the measurement of social class is also noted in other disciplines. A recent report by a task force commissioned by the American Psychological Association (Saegert et al., 2006) recommended that the APA establish a continuing Committee on Socioeconomic Status to ensure that issues of SES receive more attention by APA members, calling for more research in the conceptualization and measurement of SES.

The current study is a response to this call. We first review the literature on the conceptualization and measurement of social class or socioeconomic status, discussing the challenges in measuring this construct and in classifying consumers into social strata. Then, we propose a modeling framework for the definition and classification of social class that is grounded in theory, sufficiently flexible to handle multiple indicators of socioeconomic status using different measurement scales, and robust to missing data, a prevalent problem in survey research. We apply this measurement framework to data from Brazil in 2003 and 2009, covering a period of dramatic socioeconomic shifts in that country, and use this stratification of Brazilian households to study shifts in consumption during this period and to study the impact of socioeconomic status on consumption priorities.

2. The conceptualization and measurement of social class and socioeconomic status

2.1. Social class

Sociologists developed the concept and operationalization of social class based on the classification of occupations defining a society’s labor markets and production systems. Wright (1985) started with Karl Marx's classic division of the means of production between the bourgeoisie and proletariat (Marx & Engels, 1848), and added intermediate categories for the labor force based on their credentials and status to identify twelve occupations that typify classes within a society in terms of their relationship to the means of production. Goldthorpe (1987) combined Marx’s perspective with Weber’s (1949), further aggregating occupations according to market conditions (wage rates, upward mobility and economic stability) and working conditions (production control, authority) into seven social classes. Classification of individuals on the basis of occupation is common in Europe (Avendano, Kunst, van Lenthe et al., 2005; Griffin, Fuhrer, Stansfeld, & Marmot, 2002; Marmot et al., 1997). As we reviewed earlier, this focus on social class based on occupation (along with education, and sometimes income) is typical of social stratification schemes used by European marketers.

2.2. Socioeconomic status

While occupation is generally agreed to play an important role in an individual’s social status and in understanding socioeconomic inequalities within a society, it is also widely acknowledged that there are substantial disparities in the prestige and economic return within some occupational categories in modern societies. Rather than focusing on the role of labor and capital in production systems to identify social classes, the notion of socioeconomic status (SES) emphasizes status achievement, using education and income as the cause and effect of occupational status, respectively. The main argument behind this conceptualization of SES is that education qualifies the individual for occupations in modern societies and income is the consequence of occupational status. Education is also typically completed early in adulthood, thereby serving as an early indicator of SES that is valid for most adults.

Earlier attempts to rank individuals in terms of socioeconomic status related education and income to occupational status, producing an index (e.g., the Socio Economic Index by Duncan, 1961; the International Socio-economic Index by Ganneboom, de Graaf, & Treiman, 1992) used to rank individuals within a society (Duncan, 1961;
Hauser & Warren, 1997). However, using educational attainment and income as indicators of SES has several limitations. Educational stratification in modern societies occurs not only in the access to different levels of education, but increasingly in the quality of education received at these levels. Moreover, measuring formal education does not capture on-the-job training or other career investments that differentiate between individuals with similar levels of formal schooling. These shortcomings make educational attainment measured in years of education a limited indicator of socioeconomic status. Income is also a problematic variable to measure in surveys. While income earned from stable employment is easier to recall and report, income produced by temporary work or in informal labor markets carries larger reporting errors (Hentschel & Lanjow, 1998). Moreover, income produced from capital is more likely to be under-reported in surveys. Current household income may also be a poor indicator of the standard of living of retired individuals because it does not reflect available financial resources and disregards the cumulative effects of a lifetime of privilege or deprivation (Duncan et al., 2002). Moreover, while educational attainment reflects the individual’s potential for social status, current income is more reflective of the individual’s current condition rather than of his/her enduring or potential place in society. Friedman (1957) makes a distinction between income and wealth, considering income as having two components: transitory income and permanent income “reflecting the effect of those factors that the unit regards as determining its capital value or wealth,” (Friedman, 1957, pp. 21). Friedman argues that consumption behavior is primarily determined by permanent income, and that consumption typically correlates poorly with income because total income is a poor measure of permanent income, as people smooth their consumption over time by borrowing or drawing on savings in times of low income and investing/saving in times of high income. Consequently, wealth can vary dramatically across different social groups with similar incomes (Braverman et al., 2005).

While European marketers typically base their social stratification on the concept of social class, using occupation as the primary indicator (Urquijo & Lobl, 2003; Ware & Dinning, 2003), marketers in the “new world” base their stratification on the concept of socioeconomic status (Corrales et al., 2006), using permanent income as its latent measure. However, SES and its latent source (permanent income) are theoretical constructs that are not directly observable and must therefore be inferred from proxy variables. A common set of proxies for inferring wealth in addition to current income is the ownership of durable consumer goods and the employment of domestic labor, which provide valuable insights into how households utilize their wealth (Filmer & Pritchett, 2001; Montgomery, Gragnolati, Burke, & Paredes, 2000). These proxies serve as valuable indicators of the household’s long-run economic status; asset ownership is unlikely to change in response to short-term economic shocks and will therefore lead to a more stable measure of long-term socioeconomic status, and it is also generally measured with less error (McKenzie, 2003; Onwujeke, Hanson, & Fox-Rushby, 2006). Asset ownership is also less likely to be affected by household size and composition than consumption needs and patterns because households are likely to adjust their consumption patterns in response to economic shocks but less likely to sell assets (Filmer & Pritchett, 2001).

Another common indicator used to infer socioeconomic status is the individual’s access to public services. However, access to certain public services (paved roads, sewage systems, piped water) might be supply-restricted, thereby reflecting the regional availability of these services rather than the individual’s access to them. On the other hand, at equilibrium, these supply restrictions might be a reflection of the lack of political clout or geographic mobility and, therefore, lead to lower social status in rural areas.

A review of the extensive literature on the conceptualization and measurement of social class and socioeconomic status across multiple disciplines, briefly summarized above, shows that theory stipulates SES as a latent construct based on social resources such as income, education, occupation, access to public services, and assets accumulated through life. Income is a necessary but insufficient indicator of SES, which must be inferred from indicators that reveal the individual’s ability to move or to remain in his/her current status and his/her ability to take advantage of society’s resources.

2.3. Measurement methodology

Recently, social scientists have proposed several methodological improvements to the identification of socioeconomic classes and to the classification of individuals in to these social strata. Filmer and Pritchett (2001) apply principal components analysis to dummy-coded data on housing characteristics and the ownership of consumer durables, constructing a uni-dimensional asset index (i.e., the first principal component) as a proxy for wealth or permanent income. Vyas and Kumaranayake (2006) discuss issues related to the choice of assets to measure SES and the preparation of data for this purpose, pointing out that substantial variance (i.e., information) is ignored when only the first principal component is retained. While the use of principal components analysis of asset-ownership has been a popular approach for SES measurement via wealth indices (Filmer & Pritchett, 2001; McKenzie, 2003; Sahm & Stiefel, 2003; Vyas & Kumaranayake, 2006), it has been also the subject of criticisms from several scholars who do not disagree with the notion of using asset-ownership as proxies for wealth or permanent income, but object to the application of principal components analysis to these data on methodological grounds. For example, Howe, Hargreaves, and Huttly (2008) note that the goal of measuring wealth along a single dimension is not necessarily compatible with the underlying assumptions of principal components analysis, which is likely to produce multiple dimensions that are subsequently ignored when focusing solely on the first dimension. They also point out that the binary nature of asset-ownership data is not consistent with PCA, which assumes continuous indicators and produces the wealth index as a linear combination of these continuous variables. Kolenikov and Angeles (2009) propose the use of polychoric correlations to deal with the binary nature of asset-ownership and demonstrate that proper treatment of the indicator variables can affect the predictive validity of the resulting wealth indices. May (2006) avoids these problems by applying a unidimensional latent-trait (e.g., item response theory) model where the ownership of each asset serves as a binary indicator of the continuous, unidimensional latent trait representing the inferred wealth index. This approach is superior to PCA in its treatment of binary data and robustness to missing data, but it carries the same caveat as Filmer and Pritchett (2001) of imposing a unidimensional structure to asset-ownership. Clearly, a multidimensional latent-trait model (Reckase, 2007) could be applied to the same indicators, but one would have to wrestle with the interpretation of the multiple dimensions and with their connection to the underlying theoretical concept of permanent income. Bollen, Glanville, and Stecklov (2006) develop a structural equation model that posits a latent variable measuring the theoretical construct of “permanent income,” utilizing asset-ownership, education, occupation and income as indicators of permanent income as a formative construct.

All these attempts to build wealth indices as measurements of permanent income treat SES as a continuum, regardless of whether they are based on principal components, item-response theory or structural equation modeling, with each individual occupying a place along this unidimensional line. This notion of a SES continuum contradicts the original conceptualization of social classes as discrete categories defining the status of large sectors of a society. In other words, recent work in the measurement of SES has departed from

---

1 A more flexible unidimensional ordering of households may be attained with Mokken scale analysis (Van Schuur, 2003), which is based on the same principles of the Item Response Theory, but with a more flexible, nonparametric measurement model.
the general conceptualization of class, which is generally ignored by wealth indices (Bollen, Glanville, & Stecklov, 2001). When one reads a press report that major corporations are turning their attention to the “middle-class” in emerging economies, the reference is to a group or class of consumers rather than to the consumers’ individual placement along a continuous line. Marketers in emerging economies plan and implement their strategies at the segment level, and therefore feel more comfortable operating with classes than with individual rankings or continuous measurements.

The social-stratification framework we propose and implement next is an attempt to confront these differences between the continuous notion of SES and the discrete conceptualization of social classes. Our main objective with this framework is to define social strata on the basis of the more modern construct of permanent income but identify latent classes rather than a unidimensional continuum, because our main purpose is to produce a classification tool for market segmentation purposes in the spirit of the many marketing applications we reviewed earlier on. We apply a monotonically constrained latent class model for the identification and classification of socioeconomic strata based on indicators of permanent income, as well as social class. As we demonstrate later, this monotonically-constrained latent class model for socioeconomic stratification shares some of the features of item response theory, in the sense that it is robust to missing data and provides diagnostics about the information contained in each indicator and its meaning for the individual’s socioeconomic standing. However, rather than placing individuals on a unidimensional continuum or adding the complexity of mapping households on multiple dimensions (which might not be theoretically justified), interpreting these dimensions, and then

Table 1

Socioeconomic profiles from the proposed stratification.

<table>
<thead>
<tr>
<th>Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>7%</td>
<td>13%</td>
<td>16%</td>
<td>22%</td>
<td>21%</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Head education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete elementary</td>
<td>1%</td>
<td>1%</td>
<td>14%</td>
<td>14%</td>
<td>33%</td>
<td>43%</td>
<td>62%</td>
<td>75%</td>
<td>31%</td>
</tr>
<tr>
<td>Incomplete middle school</td>
<td>5%</td>
<td>5%</td>
<td>25%</td>
<td>25%</td>
<td>34%</td>
<td>35%</td>
<td>29%</td>
<td>22%</td>
<td>27%</td>
</tr>
<tr>
<td>Incomplete high school</td>
<td>9%</td>
<td>9%</td>
<td>18%</td>
<td>18%</td>
<td>14%</td>
<td>11%</td>
<td>6%</td>
<td>3%</td>
<td>12%</td>
</tr>
<tr>
<td>Incomplete college</td>
<td>48%</td>
<td>48%</td>
<td>34%</td>
<td>34%</td>
<td>16%</td>
<td>10%</td>
<td>3%</td>
<td>1%</td>
<td>21%</td>
</tr>
<tr>
<td>College or graduate</td>
<td>37%</td>
<td>37%</td>
<td>10%</td>
<td>10%</td>
<td>3%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Head occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private employment</td>
<td>26%</td>
<td>28%</td>
<td>38%</td>
<td>40%</td>
<td>37%</td>
<td>33%</td>
<td>25%</td>
<td>16%</td>
<td>33%</td>
</tr>
<tr>
<td>Public employment</td>
<td>17%</td>
<td>24%</td>
<td>10%</td>
<td>12%</td>
<td>8%</td>
<td>4%</td>
<td>3%</td>
<td>2%</td>
<td>9%</td>
</tr>
<tr>
<td>Employer</td>
<td>16%</td>
<td>12%</td>
<td>6%</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>Temporary rural worker</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>6%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>Other employment</td>
<td>1%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>6%</td>
<td>9%</td>
<td>9%</td>
<td>12%</td>
<td>14%</td>
</tr>
<tr>
<td>Retired</td>
<td>18%</td>
<td>18%</td>
<td>16%</td>
<td>23%</td>
<td>13%</td>
<td>24%</td>
<td>10%</td>
<td>13%</td>
<td>18%</td>
</tr>
<tr>
<td>Unemployed</td>
<td>4%</td>
<td>2%</td>
<td>4%</td>
<td>4%</td>
<td>5%</td>
<td>8%</td>
<td>4%</td>
<td>6%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>log(per capita income)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City network</td>
<td>84%</td>
<td>84%</td>
<td>74%</td>
<td>74%</td>
<td>45%</td>
<td>45%</td>
<td>2%</td>
<td>1%</td>
<td>51%</td>
</tr>
<tr>
<td>Septic tank</td>
<td>12%</td>
<td>12%</td>
<td>15%</td>
<td>15%</td>
<td>22%</td>
<td>22%</td>
<td>12%</td>
<td>3%</td>
<td>7%</td>
</tr>
<tr>
<td>Tank</td>
<td>4%</td>
<td>4%</td>
<td>9%</td>
<td>9%</td>
<td>28%</td>
<td>28%</td>
<td>55%</td>
<td>32%</td>
<td>22%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>6%</td>
<td>6%</td>
<td>3%</td>
<td>2%</td>
<td>10%</td>
</tr>
<tr>
<td>City water</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
<td>93%</td>
<td>93%</td>
<td>51%</td>
<td>29%</td>
<td>87%</td>
</tr>
<tr>
<td>Pavement</td>
<td>95%</td>
<td>95%</td>
<td>88%</td>
<td>88%</td>
<td>63%</td>
<td>63%</td>
<td>12%</td>
<td>12%</td>
<td>66%</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>2.7</td>
<td>2.15</td>
<td>1.25</td>
<td>1.25</td>
<td>1.0</td>
<td>1.0</td>
<td>0.84</td>
<td>0.49</td>
<td>1.26</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>2.86</td>
<td>2.09</td>
<td>1.53</td>
<td>1.27</td>
<td>1.13</td>
<td>1.04</td>
<td>0.84</td>
<td>0.49</td>
<td>1.26</td>
</tr>
<tr>
<td>Stoves</td>
<td>1.25</td>
<td>1.11</td>
<td>1.11</td>
<td>1.08</td>
<td>1.03</td>
<td>1.03</td>
<td>0.73</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>Freezers</td>
<td>0.67</td>
<td>0.39</td>
<td>0.12</td>
<td>0.10</td>
<td>0.06</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>Refrigerators</td>
<td>1.21</td>
<td>0.74</td>
<td>0.64</td>
<td>0.29</td>
<td>0.24</td>
<td>0.06</td>
<td>0.06</td>
<td>0.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Blenders</td>
<td>1.26</td>
<td>1.06</td>
<td>1.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.85</td>
<td>0.83</td>
<td>0.13</td>
</tr>
<tr>
<td>Vacuum cleaners</td>
<td>0.73</td>
<td>0.48</td>
<td>0.27</td>
<td>0.09</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Press irons</td>
<td>1.53</td>
<td>1.14</td>
<td>1.14</td>
<td>0.97</td>
<td>0.97</td>
<td>0.67</td>
<td>0.60</td>
<td>0.09</td>
<td>0.88</td>
</tr>
<tr>
<td>Clothes washers</td>
<td>1.04</td>
<td>0.92</td>
<td>0.83</td>
<td>0.56</td>
<td>0.56</td>
<td>0.38</td>
<td>0.11</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>Color TVs</td>
<td>3.21</td>
<td>2.03</td>
<td>2.01</td>
<td>1.29</td>
<td>1.24</td>
<td>0.88</td>
<td>0.88</td>
<td>0.22</td>
<td>1.32</td>
</tr>
<tr>
<td>Radios</td>
<td>0.97</td>
<td>0.61</td>
<td>0.59</td>
<td>0.37</td>
<td>0.37</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.43</td>
</tr>
<tr>
<td>Air conditioners</td>
<td>0.94</td>
<td>0.49</td>
<td>0.08</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Fans</td>
<td>2.38</td>
<td>1.52</td>
<td>1.66</td>
<td>0.89</td>
<td>1.10</td>
<td>0.52</td>
<td>0.48</td>
<td>0.08</td>
<td>0.98</td>
</tr>
<tr>
<td>Sewing machines</td>
<td>0.51</td>
<td>0.37</td>
<td>0.37</td>
<td>0.26</td>
<td>0.26</td>
<td>0.13</td>
<td>0.13</td>
<td>0.07</td>
<td>0.24</td>
</tr>
<tr>
<td>Water filters</td>
<td>0.42</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.33</td>
<td>0.33</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td>Automobiles</td>
<td>1.68</td>
<td>1.11</td>
<td>0.75</td>
<td>0.38</td>
<td>0.26</td>
<td>0.05</td>
<td>0.05</td>
<td>0.02</td>
<td>0.39</td>
</tr>
<tr>
<td>Bicycles</td>
<td>1.18</td>
<td>0.46</td>
<td>0.97</td>
<td>0.26</td>
<td>0.26</td>
<td>0.04</td>
<td>0.04</td>
<td>0.15</td>
<td>0.79</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>0.18</td>
<td>0.08</td>
<td>0.20</td>
<td>0.09</td>
<td>0.17</td>
<td>0.04</td>
<td>0.15</td>
<td>0.06</td>
<td>0.12</td>
</tr>
<tr>
<td>Personal computers</td>
<td>1.33</td>
<td>0.86</td>
<td>0.52</td>
<td>0.26</td>
<td>0.26</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>Water purifiers</td>
<td>0.23</td>
<td>0.17</td>
<td>0.08</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Microwave ovens</td>
<td>0.96</td>
<td>0.74</td>
<td>0.52</td>
<td>0.33</td>
<td>0.33</td>
<td>0.07</td>
<td>0.01</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>Parabolic antennas</td>
<td>0.24</td>
<td>0.15</td>
<td>0.33</td>
<td>0.11</td>
<td>0.11</td>
<td>0.08</td>
<td>0.04</td>
<td>0.05</td>
<td>0.22</td>
</tr>
<tr>
<td>DVD players</td>
<td>1.71</td>
<td>1.04</td>
<td>0.94</td>
<td>0.60</td>
<td>0.60</td>
<td>0.53</td>
<td>0.22</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td>Clothes dryers</td>
<td>0.28</td>
<td>0.12</td>
<td>0.10</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Mixers</td>
<td>0.95</td>
<td>0.72</td>
<td>0.72</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.40</td>
<td>0.35</td>
</tr>
<tr>
<td>Hair dryers</td>
<td>1.19</td>
<td>0.73</td>
<td>0.59</td>
<td>0.28</td>
<td>0.17</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.27</td>
</tr>
<tr>
<td>Dishwashers</td>
<td>0.33</td>
<td>0.15</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Variables in bold were constrained to vary monotonically across strata.
hierarchically categorizing the households into classes, our framework directly identifies stratified socioeconomic classes and classifies individuals into these strata.

### 3. Socioeconomic stratification in Brazil

When compared to the trajectory of some of the other “BRICS” economies such as China and India, Brazil’s growth rate in the first decade of the 21st century is not as impressive, at 2.9% annual growth in per capita GDP between 2003 and 2009. The most remarkable aspect of this growth in the Brazilian economy is that it happened with income redistribution rather than concentration. In the first decade of the 21st century, the Gini coefficient in Brazil declined from its peak of 0.609 in 1990 to 0.545 in 2009 (Neri, 2010), which combined with economic growth led to a 45% reduction in the population living below the poverty level between 2003 and 2009 (Neri, 2010) and to the emergence of a “new middle class” in Brazil. However, when one mentions the “middle class” within a society it is often unclear exactly what this cohort represents. The main purpose of the model and application we present next is to develop an objective socioeconomic stratification that can be applied across studies within the same society, thereby providing a consistent definition of socioeconomic status that can also be used as a basis for comparisons within and across studies.

As briefly discussed earlier, a widely applied criterion for the classification of consumers into socioeconomic classes in Brazil is the “Criterio Brasil,” which is based on education, the number of certain durable consumer goods present in the home, the number of bathrooms, and the number of domestic workers employed by the household. This criterion was developed as a predictor of income by estimating a classic Mincerian income regression (Mincer, 1958) to define the weights for each indicator in the final income score, and then classifying individuals into eight strata based on this score (Neri, 2010), thereby leading to a classification based on predicted current income rather than on the theoretical concept of permanent income. In other words, despite the fact that “Criterio Brasil” uses asset ownership as an indicator, it is still focused on household income because its score is defined by a Mincerian income regression model predicting current household income as the dependent variable.

The framework for socioeconomic class identification and classification we propose next differs from the popular “Criterio Brasil” in several important ways. First, rather than building a predictive score for current household income and stratifying the population based on this score, we directly identify classes based on indicators of permanent income or wealth (e.g., the number of consumer durables owned, the employment of household help) as well as social class (e.g., the occupation of the head of household, educational attainment, access to public services). Second, rather than creating a unidimensional wealth index and stratifying the population based on this index, we directly identify latent classes that are ordered in terms of the indicators of social class and permanent income, thereby identifying natural strata in the population based on two closely related theoretical constructs. Third, rather than classifying households into strata on the basis of a linear score and an arbitrary set of cut-off points, our framework directly classifies each household probabilistically into “naturally” occurring strata using a monotonically-constrained latent-class model that avoids the assumption of unidimensionality and is robust to missing data. This robustness to missing data is particularly valuable for marketing researchers, because it makes it possible to compare stratifications obtained from different sets of indicators (Balasubramanian & Kamakura, 1989).

#### 3.1. A monotonically-constrained latent class model for socioeconomic stratification

Let \( Y_i \) represent a K-dimensional vector of continuous, ordinal or nominal indicators of permanent income or social standing for household \( i \) and assume we have gathered these indicators from a sample of \( N \) households. Our goal is to identify \( S \) ordinal socioeconomic strata in our sample that are ordered according to \( K \) indicators. For example, if \( y_{ik} \) is a continuous indicator, then the mean of this indicator should monotonically increase or decrease across the ordinal strata; similarly, if \( y_{ik} \) is a binary indicator, then the proportion for this binary indicator should monotonically increase or decrease across the ordinal strata. As we briefly show below, this can be easily accomplished with a monotonically-constrained latent class model (Vermunt, 2001). The likelihood function for the standard latent class model for household \( i \) is given by

\[
L(Y_i | \Theta, \Pi) = \sum_{s=1}^{S} \pi_s \prod_{k=1}^{K} p(y_{ik} | \theta_{ks}).
\]  

where \( \pi_s \) is the estimated size of socioeconomic class \( s \) in the population, \( \theta_{ks} \) collects the response parameters for item \( k \) and segment \( s \), and \( p(y_{ik} | \theta_{ks}) \) is the probability associated with the observed indicator \( y_{ik} \) given that household \( i \) belongs to socioeconomic class \( s \). The form of the probability function \( p(y_{ik} | \theta_{ks}) \) will depend on whether the \( k^{th} \) indicator is binary, categorical, ordinal or continuous. For example, if the indicator is continuous, then

\[
p(y_{ik} | \theta_{ks}) = \frac{1}{\sigma} \exp\left(-\frac{1}{2} \left( \frac{y_{ik} - \mu_{ks}}{\sigma} \right)^2 \right).
\]

On the other hand, if the \( k^{th} \) indicator is categorical with \( M \) categories, then

\[
p(y_{ik} | \theta_{ks}) = \frac{\exp(\phi_{ks}/\gamma_{ks})}{\sum_{m=1}^{M} \exp(\phi_{km}/\gamma_{ks})}.
\]

To ensure that the latent strata are ordered, we specify an ordinal latent class model that
imposes monotonic constraints on the class positional parameters \( \theta_{ks} \leq \theta_{ks+1} \) or \( \theta_{ks} \geq \theta_{ks+1} \) for all manifest variables where the researcher has a priori information on the theoretical order of these positional parameters across social strata. Estimation of the monotonically-constrained latent class model is easily attained with commercial software (Latent Gold, Vermunt & Magidson, 2010). While the model described here is basically a standard latent class model with monotonic constraints on other response parameters, it offers many valuable features for the purpose of social stratification. First, as shown in the literature, latent class analysis is robust to missing data (Kamakura & Wedel, 1997), which leads to a flexible classification framework where different researchers can categorize their samples into the same social strata, even though they might have collected different subsets of indicators. Second, the monotonic constraint in the response parameters across classes leads to ordered latent classes that are consistent with the theoretical concept of social stratification. Third, latent class analysis directly identifies naturally occurring latent classes that best explain the relationship among manifest variables, directly mapping these manifest variables into discrete social strata, thereby bypassing the arbitrary categorization of a unidimensional continuum into classes.\(^2\) Despite these valuable advantages, we are not aware of any prior work using this methodology for social stratification.

3.2. Brazilian socioeconomic classes in 2003 and 2009

To illustrate the features of our proposed framework for socioeconomic stratification and to understand the dramatic shifts in socioeconomic structure and consumption in Brazil in recent years, we apply our framework to the household characteristics gathered in 2003 and 2009 by the “Instituto Brasileiro de Geografia e Estatística” (IBGE) as part of the Survey of Household Budgets (“Pesquisa de Orçamento Familiar”, or POF), which we later relate to consumption expenditures observed in the same period. As discussed earlier, we depart from traditional wealth indices by combining indicators of permanent income such as the number of consumer durables owned, employment of household help, and dwelling characteristics, with indicators of social class such as education, the occupation of the head of household, and access to public services. Summary statistics for these indicators are listed in the last column of Table 1.

We applied the monotonically-constrained latent class model to the sample of 48,568 households in 2003 and 55,685 in 2009, using projection weights provided by IBGE to ensure that the results are directly projectable to the population of households in the country in both years. We chose to identify the same number of social strata as

---

2 A decision must be made on the desired number of classes, which can be either done with statistical information criteria such as the Consistent Akaike's Bayesian Criterion (CAIC) or the Bayesian Information Criterion (BIC), or by the researcher based on the purposes of the stratification analysis (see Wedel & Kamakura, 2000).
lump the majority of the population into very large classes, making it more informative for studies focusing on the wealthy minority than for studies focusing on the less (economically) fortunate majority, including the middle class, which has been the recent focus of marketers in Brazil. In contrast, our proposed social stratification produces a more symmetric distribution, with the bulk of the population in the central four strata (3 to 6), leading to a more balanced classification of the population in socioeconomic terms. With this more balanced distribution, our socioeconomic classification is more informative both at the middle and extreme strata. Second, because the distribution of socioeconomic strata produced by our proposed model is less skewed, it produces clearer evidence of the dramatic shifts in socioeconomic stratification between 2003 and 2009.

The top panel of Fig. 1 shows how the population of households has shifted towards the higher strata on the left, particularly the households classified in the middle to low socioeconomic strata. This upward shift has led to over 20.5 million Brazilians rising above the poverty threshold (R$145 or US$83 of monthly per capita income based on the exchange rate in December 2009) between 2003 and 2009. This emergence of a new class of economically-relevant consumers requires the repositioning of many businesses to better cater to their needs, which change as these consumers move upwards in the social stratification, with a growing interest in and consumption of categories not previously consumed, such as food away from home, cosmetics, automobiles and air travel (Souza & Lamounier, 2010).

### 3.3. Consumption concentration by category and social strata

The differences between our proposed socioeconomic stratification and the social classification based on the “Criterio Brasil” become more evident when comparing the distribution of the annual

| Table 3 | Demographic analysis of the socioeconomic strata in Brazil (2003 and 2009). |
| --- | --- | --- |
| Effect | Estimate | Std. error |
| State capital | 0.71 | 0.19 |
| Metropolitan area | 0.36 | 0.34 |
| Rural | −1.44 | 0.20 |
| Residents = 1 | −0.15 | 0.061 |
| Residents = 2 | 0.12 | 0.054 |
| Residents = 3 | 0.31 | 0.49 |
| Residents = 4 | 0.48 | 0.045 |
| Residents = 5 | 0.37 | 0.045 |
| Residents = 6 | 0.23 | 0.050 |
| Age = up to 24 | −0.22 | 0.044 |
| Age = 25 to 29 | 0.18 | 0.037 |
| Age = 30s | 0.52 | 0.032 |
| Age = 40s | 0.64 | 0.032 |
| Age = 50s | 0.46 | 0.031 |
| Age = 60s | 0.19 | 0.032 |
| Married | 0.42 | 0.021 |
| Single male | −0.33 | 0.032 |
| White | 1.91 | 0.091 |
| Black | 0.74 | 0.094 |
| Asian | 2.35 | 0.141 |
| Brown | 0.79 | 0.091 |
| Kids10_14 | 1.16 | 0.088 |
| Kids10_14 | 0.95 | 0.088 |
| Kids15_19 | 0.33 | 0.105 |
| Kids25_29 | −0.27 | 0.073 |
| Kids30_34 | −0.11 | 0.041 |
| Kids35_39 | −0.14 | 0.072 |
| Kids40_44 | −0.16 | 0.074 |
| Kids45_49 | 0.54 | 0.160 |
| [Black*][year = 2003] | −0.48 | 0.164 |
| [Asian*][year = 2003] | −0.53 | 0.230 |
| [Brown*][year = 2003] | −0.58 | 0.160 |
| [Age & 30s]*[year = 2003] | 0.16 | 0.048 |
| [Age & 40s]*[year = 2003] | 0.14 | 0.048 |
| [State capital]*[year = 2003] | 0.18 | 0.028 |
| [Metropolitan area]*[year = 2003] | 0.35 | 0.050 |
| [Rural]*[year = 2003] | −0.47 | 0.029 |

### Table 4 | Entropy classification for each indicator and socioeconomic stratum. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Expected posterior entropy</td>
<td>Expected drop in entropy</td>
</tr>
<tr>
<td>Per capita income</td>
<td>1.56</td>
<td>24.5%</td>
</tr>
<tr>
<td>Color TVs</td>
<td>1.59</td>
<td>22.1%</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>1.65</td>
<td>17.7%</td>
</tr>
<tr>
<td>Personal computers</td>
<td>1.71</td>
<td>13.6%</td>
</tr>
<tr>
<td>Sewage</td>
<td>1.71</td>
<td>13.5%</td>
</tr>
<tr>
<td>Automobiles</td>
<td>1.72</td>
<td>13.3%</td>
</tr>
<tr>
<td>Microwave ovens</td>
<td>1.72</td>
<td>12.9%</td>
</tr>
<tr>
<td>Education</td>
<td>1.72</td>
<td>12.8%</td>
</tr>
<tr>
<td>Clothes washers</td>
<td>1.72</td>
<td>12.8%</td>
</tr>
<tr>
<td>DVD players</td>
<td>1.74</td>
<td>12.0%</td>
</tr>
<tr>
<td>Food mixers</td>
<td>1.75</td>
<td>10.9%</td>
</tr>
<tr>
<td>Press iron</td>
<td>1.75</td>
<td>10.8%</td>
</tr>
<tr>
<td>Hairdryers</td>
<td>1.76</td>
<td>10.1%</td>
</tr>
<tr>
<td>Refrigerators</td>
<td>1.78</td>
<td>9.4%</td>
</tr>
<tr>
<td>Pavement</td>
<td>1.78</td>
<td>9.2%</td>
</tr>
<tr>
<td>Fans</td>
<td>1.79</td>
<td>8.8%</td>
</tr>
<tr>
<td>Blenders</td>
<td>1.79</td>
<td>8.6%</td>
</tr>
<tr>
<td>Freezers</td>
<td>1.79</td>
<td>8.4%</td>
</tr>
<tr>
<td>Vacuums</td>
<td>1.80</td>
<td>7.8%</td>
</tr>
<tr>
<td>City water</td>
<td>1.81</td>
<td>7.4%</td>
</tr>
<tr>
<td>Occupation of HH</td>
<td>1.85</td>
<td>5.0%</td>
</tr>
<tr>
<td>Stereos</td>
<td>1.85</td>
<td>4.9%</td>
</tr>
<tr>
<td>Air conditioners</td>
<td>1.86</td>
<td>4.7%</td>
</tr>
<tr>
<td>Bicycles</td>
<td>1.88</td>
<td>3.3%</td>
</tr>
<tr>
<td>Mails</td>
<td>1.88</td>
<td>3.1%</td>
</tr>
<tr>
<td>Dish antennas</td>
<td>1.89</td>
<td>2.8%</td>
</tr>
<tr>
<td>Stoves</td>
<td>1.90</td>
<td>2.5%</td>
</tr>
<tr>
<td>Dishwashers</td>
<td>1.90</td>
<td>2.4%</td>
</tr>
<tr>
<td>Clothes dryers</td>
<td>1.91</td>
<td>1.8%</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>1.91</td>
<td>1.8%</td>
</tr>
<tr>
<td>Water purifiers</td>
<td>1.91</td>
<td>1.6%</td>
</tr>
<tr>
<td>Sewing machines</td>
<td>1.92</td>
<td>1.3%</td>
</tr>
<tr>
<td>Radios</td>
<td>1.92</td>
<td>1.2%</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>1.93</td>
<td>0.9%</td>
</tr>
<tr>
<td>Water filters</td>
<td>1.94</td>
<td>0.1%</td>
</tr>
<tr>
<td>Black and white TVs</td>
<td>1.94</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Note that because we identified the latent social strata by pooling the data between 2003 and 2009, we are able to see the dramatic shifts in social stratification in Brazil during this relatively short period of six years, as shown in Fig. 1, compared to the shifts measured in terms of the existing industry criterion (“Criterio Brasil”) applied to the same data. This comparison reveals two major differences between the two stratification procedures. First, the social classes defined by “Criterio Brasil” are skewed in size towards classes C1, C2 and D, with much fewer households classified in the first three classes (A1, A2 and B). This highly skewed distribution helps to discriminate the wealthy minority from the rest of the population, but tends to

3 This decision could be guided by information criteria as discussed earlier or by cross-validation tests (Wedel & Kamakura, 2001), but ultimately should depend on the intended use of the stratification scheme. Given that the most widely utilized stratification scheme in Brazil contains eight strata, and given our intention to use this scheme as a basis for comparison, we chose to estimate our model with eight ordered classes.

in “Criterio Brasil” so that we can directly compare the two social-stratification schemes. Given the large number (over 500) of parameters involved, we do not report them here (they are available upon request); a more intuitive understanding of the socioeconomic classes can be obtained from the class profiles reported in Table 1.
consumption budget implied by the two criteria, shown in Table 2. This table shows that “Criterio Brasil” produces social classes that are uneven in terms of economic relevance to marketers, while the socioeconomic strata identified by the proposed model are more evenly distributed in terms of economic relevance, except for the two lowest strata (7 and 8), which are less economically relevant. For example, the first two classes (A1 and A2) from “Criterio Brasil” represent a very small fraction of both households (1.6%) and cumulative consumption (8%), while in contrast, our proposed stratification produces top two strata (1 and 2) containing 12.5% of the population and accounting for 35.3% of consumption. On the other extreme, the situation is reversed; “Criterio Brasil” lumps almost 40% of the households into the two bottom strata accounting for about 19% of consumption, while the proposed stratification places 22% of households in the two bottom strata, accounting for 14% of total consumption. These results demonstrate that “Criterio Brasil” provides more details about consumption among the top minority of wealthier Brazilians, while our proposed stratification produces more evenly distributed strata in terms of both size and consumption.

While the results from Fig. 1 and Table 2 show that the two stratification schemes produce distinct class distributions both in terms of population and economic relevance, the question still remains as to whether they differ in their ability to differentiate among consumers in terms of consumption. To answer this question, we analyzed consumption expenditures across our sample of 55,685 households in 2009, estimating the percentage of variance in projected consumption explained by the two stratification schemes across the Brazilian population in 2009 for each of 21 consumption categories. The results from this comparison are displayed in Fig. 2, which shows that the proposed stratification generally explains differences in consumption better than the traditional Mincerian approach. The only exceptions are clearly non-essential categories such as Jewelry and Accessories, Leisure Travel, Personal Services, Car Maintenance and Other Expenses.

These results once again demonstrate that by devoting the four top strata to the wealthiest 17% of the Brazilian population, the Mincerian approach provides more details about the wealthy minority, thereby better explaining consumption expenditures in some of the non-essential categories. By distributing the population more evenly across the eight strata, our proposed stratification explains a higher variance in expenditures for most (16) of the 21 consumption categories. At first glance, these differences in explanatory power might seem small, but two important factors must be taken into account. First, these measures of explanatory power are obtained across a large sample of more than 50,000 observations, and therefore, a strong regression towards the mean should be expected, leading to smaller (but highly significant) differences. Second, both classification schemes rely on exactly the same indicators, with the result that their classifications cannot be dramatically different; in other words, one should not expect households classified at the bottom (or

---

### Table 5

Classification comparison (full set vs. top 15 and 14 selected items).

<table>
<thead>
<tr>
<th>Stratification based on</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>top 15 indicators</td>
<td>2106</td>
<td>288</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 14 indicators</td>
<td>1414</td>
<td>1592</td>
<td>8434</td>
<td>1224</td>
<td>211</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10747</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 13 indicators</td>
<td>2</td>
<td>233</td>
<td>1519</td>
<td>8482</td>
<td>1594</td>
<td>239</td>
<td>2</td>
<td>0</td>
<td>12071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 12 indicators</td>
<td>0</td>
<td>678</td>
<td>678</td>
<td>1917</td>
<td>20633</td>
<td>2698</td>
<td>563</td>
<td>0</td>
<td>26489</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 11 indicators</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>323</td>
<td>3566</td>
<td>18168</td>
<td>1992</td>
<td>337</td>
<td>24387</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 10 indicators</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>669</td>
<td>2479</td>
<td>9784</td>
<td>752</td>
<td>18885</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>top 9 indicators</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>342</td>
<td>885</td>
<td>9711</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Table 6

Cumulative consumption expenditures by socio economic strata for selected categories.

<table>
<thead>
<tr>
<th>Year</th>
<th>Social class</th>
<th>2003</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Social class</td>
<td>2003</td>
<td>2009</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
top) according to one scheme to move to the top (or bottom) according to the other scheme, and therefore, the differences in stratification straddle neighboring strata. Most importantly, our proposed stratification produces consistently better explanation of consumption, with only a few (5) exceptions.

3.4. Socioeconomic stratification and demography

To understand these social strata and the movements between 2003 and 2009 in demographic terms, we applied an ordinal logistic regression to the individual classifications of the 104,253 households in our sample, using their demographic profiles and year of data collection as predictors. The results are reported in Table 3, where social strata are coded so that a positive coefficient reflects a propensity to belong to higher strata (i.e., higher permanent income); to simplify presentation, only the statistically significant parameters (p<0.01) are reported. Based on these results, we conclude that:

- Households living in state capitals and other metropolitan areas are more likely to be in higher socioeconomic strata, while households living in rural areas tend to be in lower strata.
- Mid-size households (three to five members) are more likely to be in higher strata than smaller or larger households.
- Households with middle-aged heads (30s to 50s) are more likely to be in higher strata than those with younger or older heads.
- Households headed by couples are more likely to be in higher strata, while those headed by single males are more likely to be in lower strata.
- Households headed by Asians are more likely to be in the upper strata, followed by Whites and Browns.
- Households without children under 19 years of age or with children over 25 are more likely to be in the upper strata.
- The shifts observed between 2003 and 2009 were not evenly distributed across demographic groups.
- The position of households headed by Whites, Browns, Asians and Blacks improved between 2003 and 2009.
- The position of households with heads in the 30s and 40s declined between 2003 and 2009.
- The position of households living in state capitals and other metropolitan areas declined between 2003 and 2009.

3.5. Useful features of the proposed socioeconomic stratification framework

Because the socioeconomic classes were identified on the basis of a monotonically-constrained latent class model, classification of individual households into the strata is robust to missing data (Balasubramanian & Kamakura, 1989; Wedel & Kamakura, 2001). For example, if only a subset of $K’$ (where $K’<K$) indicators $Y^k_i = \{y_{ik}; k=1,...,K’\}$ is observed for household $i$, the posterior probability that the household belongs to stratum $s$ is obtained from:

$$\Pr\{i\in s|Y^k_i\} = \frac{\pi_s \prod_{k=1}^{K’} p(y_{ik}|\theta_{ik})}{\sum_s \pi_s \prod_{k=1}^{K’} p(y_{ik}|\theta_{ik})}. \quad (2)$$

In other words, a household can be classified into the socioeconomic strata using all the available information, even when the number of indicators available from that household is limited relative to the set used to identify and define the socioeconomic strata. In contrast, social stratification based on traditional wealth indices can only be done when all variables composing the index are measured.

The contribution of each individual indicator $k$ to socioeconomic stratification can be measured by the expected posterior classification entropy

$$E_k = \int \sum_i \left[ -\tau_{ks} \log(\tau_{ks}) - (1-\tau_{ks}) \log(1-\tau_{ks}) \right] f(y_k) dy_k, \quad (3)$$

where $\tau_{ks} = \frac{\pi_s p(y_{ik}|\theta_{ik})}{\sum_s \pi_s p(y_{ik}|\theta_{ik})}$ and $f(y_k)$ is the observed frequency distribution of the indicator ($y_k$).

This entropy measure, shown in Table 4, indicates the classification information that can be expected from each indicator variable, and therefore can be used to select indicators that provide the best classification of households into the social strata. The second column of Table 4 displays the expected posterior classification entropy for each item, computed as shown above, while the third column shows the improvement (drop) in entropy relative to the entropy based on prior membership probabilities (class sizes). These statistics show that the most informative items for social stratification are per capita income (in log); the numbers of color TVs, bathrooms, personal computers, automobiles and microwave ovens; access to sewage
treatment; and the educational attainment of the head of the household.

Taking advantage of the robustness of latent class analysis to missing data, and of the entropy measure for each social indicator in Table 4, a subset of items can be identified for a more parsimonious measurement of socioeconomic status that will not require too much time and effort, making it suitable for a consumer survey. We illustrate and test this feature by selecting a subset of indicators and comparing the classification based on this subset with the classification from the full sample. For this test, we used the top 15 of the 36 indicators listed in Table 4, as well as a similar set excluding per capita income, which is often difficult to measure in a survey (Hentschel & Lanjow, 1998). The social stratification obtained with these two subsets is compared to the original stratification (based on the calibrated model on all items) in Table 5. Socioeconomic stratification based on the top 15 indicators matched the classification obtained with all 36 indicators in 76.3% of the 104,659 households, while ignoring per capita income led to a match in 72.3% of the cases. These hit rates are quite reasonable, considering the very large sample size involved. Most importantly, Table 5 shows that the mismatched cases relative to the full-set stratification are mostly assigned to neighboring strata; when the nearest strata are included as correct matches, the matching rates increase to 96.0% and 94.8%, respectively.

The empirical evidence from Table 5 discussed above has important implications for marketing researchers, as it demonstrates that social stratification based on the proposed model is fairly robust to missing data, allowing researchers to compare social stratification based on different indicators. In other words, consumers from two different studies can be classified into the same social strata, even if the studies gathered different classification data for social stratification. This feature of the proposed stratification framework is particularly useful to marketing researchers who often need to relate results from different sources, each using different criteria for socioeconomic stratification, for example when concatenating results from media research, consumer research, internal sales reports and other sources. As long as these different sources use subsets of the items listed in Table 4, even if these items are not the same, our proposed framework makes it possible to aggregate the results from these different sources into compatible socioeconomic segments. Marketing researchers may take advantage of this to produce comparable social stratification schemes based on different indicators or to concatenate results obtained from different sources, using socio-economic stratification as the common basis for relating these independent sources. Our empirical results also show that substantial savings can be attained in data collection (asking only 14 out of 36 questions) with little loss in the accuracy of social stratification. This robustness to missing data is particularly important in survey research where respondents often decline to answer questions related to income, educational attainment and occupation, which form the basis for many socioeconomic stratification schemes (Corrales et al., 2006; Thompson & Hickey, 2005). The ability to select a small subset of the indicators of permanent income and social class to stratify a sample leads to shorter surveys, with lower costs and better data quality due to better response rates.

In conclusion, the conceptual discussion and empirical results presented so far demonstrate that the proposed framework for social stratification offers several benefits over the traditional Mincerian approach. First, rather than attempting to predict current income through the estimation of a Mincerian income regression and arbitrarily stratifying the population based on predicted current income, our framework is based on the theoretical concepts of social class and socioeconomic status, using multiple indicators of permanent income and social class to identify naturally occurring (latent) classes that are ordered in terms of these manifest indicators. Second, when applied to the household population of Brazil in 2003 and 2009, the proposed framework produces a more even stratification than the prevalent Mincerian approach, which provides finer details among the wealthy minority than among the majority, particularly the middle strata, which have been the recent focus of attention by most marketers in emerging markets. Third, because it produces more evenly distributed strata not only in size, but also in income, the proposed framework generally explains consumption better than the Mincerian approach, except for non-essential consumption, where attention to the wealthy minority is warranted. Fourth, we provide empirical evidence that by utilizing a latent class model, our framework is robust to missing data, allowing marketing researchers to produce a social stratification that is comparable across studies, even when these studies utilize different measurement indicators, which is quite useful when comparing marketing research results obtained from different sources.

4. Socio-economic stratification and consumption in Brazil

The link between socioeconomic status and consumption is almost definitional, with some scholars (e.g., Deaton, 1992) advocating the use of consumption expenditures as indicators of SES. This explains why the categorization of households into socioeconomic strata is commonly used by marketers as a market segmentation tool, particularly in emerging economies. In this second section of our study, we relate socioeconomic stratification in Brazil in 2003 and 2009 to consumer expenditures to demonstrate that in many expenditure categories, the myopic focus on the wealthy minority in emerging economies is unwise because it ignores a substantial portion of the market that, at least in Brazil, has exhibited the fastest growth with economic development. For this purpose we use expenditure data gathered by the same source (IBGE/POF) from the same households we used to develop, implement and test our socioeconomic stratification model. Aside from measuring the demographic profile and performing a complete inventory of the households’ dwelling and physical assets, the IBGE/POF program also gathers detailed data on the weekly and monthly expenditures of each household, which are annualized to reflect each household’s annual consumption budget. This wealth of data allowed us to directly relate consumption expenditure to the demographic profile and socioeconomic status of the 104,353 households in our sample, which can be projected to the country’s population based on the sample projection weights provided by the data source.

Table 6 shows how consumption expenditures are concentrated among the socio-economic strata across a few selected product/service categories. The overall conclusion from this table is that even though Brazil still has one of the highest income concentrations in the world, with a Gini coefficient of 0.545 in 2009 (Neri, 2010), the concentration in expenditures across socioeconomic strata varies substantially and is reasonably low for the more essential consumption categories, as measured by their respective Gini coefficients across socioeconomic strata. For example, expenditures in essential consumption categories such as Grains, Starches, Oils and Fats, Sugars, Animal Protein, and Urban Transportation are more evenly distributed, with Gini coefficients in the 0.02 to 0.28 range. In contrast, non-essentials such as Private Education, Private Health Insurance, Vacation Home, Personal and Professional Services, show highly concentrated expenditures, with Gini coefficients in the 0.44 to 0.68 range. A marketer competing in dairy, bakery, beverages, hygiene and personal care, or cleaning products would be unwise to focus on the top three socioeconomic strata (representing 28% of all households in 2009), because this narrow targeting would miss more than half of the market value in these categories in 2009. This empirical evidence suggests that the policy of “fighting from high ground” with premium brands followed by some global corporations competing in the packaged-goods industry might work well in developed economies, but might not transfer well to emerging markets such as Brazil for
these high-penetration product categories. In these product categories and markets, where the middle and lower socioeconomic strata represent a substantial and growing portion of total demand, it might be better to pursue a strategy that is focused on “good-enough products” (Bach, 1997; Capps, 2009) appealing to the budget-constrained, price-sensitive masses. The only consumption categories where it matters is the highly concentrated categories such as Travel or Recreation and Culture. Another way of studying consumption expenditures across socioeconomic strata is through the comparison of Engel curves (Aitchison & Brown, 1954; Du & Kamakura, 2008) relating expenditure shares for each consumption category to the total budget, which we classify in deciles (1 represents the lowest budget decile). Fig. 3 shows the Engel curves for two consumption categories as examples.4 The left panel of Fig. 3 shows that the Engel curves for Cleaning Products increase with budget (towards the higher deciles) within each stratum, the typical pattern for non-essential goods/services. However, for the same budget decile, the allocated shares are higher for the lower strata than for the upper socioeconomic strata. In other words, the total consumption budget has a positive effect on the allocated shares in this category, but households in the upper classes have a lower priority for goods in this category. In contrast, the patterns for Housing within each stratum are more typical of essential goods/services: as the budget increases (higher budget deciles), expenditure shares in these categories decrease. However, for the same consumption budget (same decile), the Engel curves shift upwards as we move up in the social ladder (from lower to upper strata). In the Housing case, the budget effect is negative (the share allocated to the category decreases with larger budgets) within each stratum, but these categories are higher priorities for households in the upper strata.

The patterns shown in Fig. 3 and discussed above suggest two different effects of socioeconomic status on consumption across budget categories: 1) a budget effect, showing changes in the slope of the Engel curve relative to the total budget, and 2) a shift in consumption priorities showing upward or downward shifts in allocated shares for the same budget, depending on the socioeconomic stratum. Therefore, socioeconomic status affects consumption expenditures in two ways. First, wealthier households in the upper strata have larger consumption budgets and therefore can devote a higher share of those budgets to less essential consumption categories such as Jewelry and Accessories than households in lower strata. Second, socioeconomic status may also affect consumption priorities so that households from different strata with similar consumption budgets will allocate different shares of their budgets to the same consumption category. The first (budget) effect is a reflection of resource constraints; households with a more limited budget must first pay for the essential elements of daily life, leaving less of the budget available for luxuries, while households with similar priorities but larger budgets can afford to spend more of their budgets on the luxuries. The second effect (differences in consumption priorities) is a reflection of sociocultural differences across strata; under similar resource constraints,
households in different strata will allocate the same budget differently across consumption categories because they assign different priorities to the competing consumption categories. Some social scientists (e.g., Bourdieu, 1984) attribute these differences in consumption priorities to class distinctions in “taste,” while others (Hirsch, 1976; Schor, 1999) argue that conspicuous consumption (on non-essential goods) has signaling value because consumers in the higher strata draw value not only from the actual consumption of luxury goods, but also from sending the signal that they can afford to spend more than other consumers (Kamakura & Du, 2012).

4.1. Consumption budget allocation model

To capture and isolate these two distinct effects on budget allocations and better assess the contribution of socioeconomic stratification in explaining consumer expenditures, we fit a consumption budget allocation model using category prices, household composition and socioeconomic class as predictors. The model we used is a simple variant of the budget allocation model recently applied by Du and Kamakura (2008) in a study of expenditure patterns among American households. In particular, our model incorporates observed heterogeneity, captured by indicators of family composition and the classification of each household into a socioeconomic stratum. Following Du and Kamakura (2008), we assume that household \( h \) maximizes a continuously differentiable quasi-concave direct utility function \( G(x_h) \) over a set of \( J \) non-negative quantities \( x_h = (x_{h1}, x_{h2}, \ldots, x_{hJ}) \), subject to a budget constraint \( p'x_h \leq m_h \), where \( p = (p_1, p_2, \ldots, p_J) > 0 \), \( p_i \) is the price of good \( i \), and \( m_h \) is household \( h \)'s total consumption budget. We use the Stone–Geary utility function, which has the form:

\[
G(x_h) = \sum_{i=1}^{J} \alpha_{ih} \ln(x_{ih} - \beta_{ih})
\]

where \( \alpha_{ih} > 0 \), \( (x_{ih} - \beta_{ih}) > 0 \), and \( J \) is the number of all available consumption categories. Note the \( h \) subscript in \( \alpha_{ih} \), which implies that all marginal utilities are household-specific. This budget allocation problem implies that the household incrementally allocates its disposable income to the consumption category that produces the highest marginal utility per dollar, given the current expenditure levels \( x_h \) until the budget limit is reached, or \( \sum_{i=1}^{J} p_i x_{ih} = m_h \). Solving this optimization problem leads to an expenditure system that is linear in total budget and prices,

\[
p_i X_{ih} = p_i \beta_{ih} + \theta_{ih} \left( m_h - \sum_{j=1}^{J} p_j \beta_{jih} \right), \quad i = 1, 2, \ldots, J
\]

with expected allocated shares given by

\[
s_{ih} = \frac{p_i \beta_{ih}}{m_h} + \theta_{ih} \left( 1 - \sum_{j=1}^{J} \frac{p_j \beta_{jih}}{m_h} \right).
\]
Counterfactual simulations comparing budget allocations under different scenarios.

<table>
<thead>
<tr>
<th>Consumption Category</th>
<th>Allocating the actual budget</th>
<th>Allocating the median budget</th>
<th>Allocating a budget 10% larger than actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expenditure shares</td>
<td>Expenditure shares</td>
<td>Expenditure shares</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>Middle</td>
<td>Lower</td>
</tr>
<tr>
<td>Food bev</td>
<td>16.6%</td>
<td>22.1%</td>
<td>29.6%</td>
</tr>
<tr>
<td>Food bevOut</td>
<td>5.5%</td>
<td>4.9%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Utilities</td>
<td>3.6%</td>
<td>5.6%</td>
<td>7.2%</td>
</tr>
<tr>
<td>Home maintenance</td>
<td>0.6%</td>
<td>0.3%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Cleaning prods</td>
<td>2.7%</td>
<td>2.8%</td>
<td>2.8%</td>
</tr>
<tr>
<td>Home furnishings</td>
<td>1.7%</td>
<td>1.8%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Communications</td>
<td>7.3%</td>
<td>6.4%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Public transport</td>
<td>5.0%</td>
<td>6.8%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Car maintenance</td>
<td>11.3%</td>
<td>7.7%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Travel</td>
<td>1.8%</td>
<td>1.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Apparel</td>
<td>5.6%</td>
<td>6.5%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Jewelry accessories</td>
<td>0.4%</td>
<td>0.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Hygiene care</td>
<td>3.6%</td>
<td>4.2%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Health</td>
<td>3.2%</td>
<td>3.9%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Health insurance</td>
<td>5.5%</td>
<td>2.4%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Education</td>
<td>1.5%</td>
<td>0.7%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Recreation cul</td>
<td>2.7%</td>
<td>2.1%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Prof services</td>
<td>0.7%</td>
<td>0.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Personal serv</td>
<td>2.7%</td>
<td>1.1%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Other expenses</td>
<td>3.4%</td>
<td>2.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Housing</td>
<td>14.4%</td>
<td>16.5%</td>
<td>20.1%</td>
</tr>
</tbody>
</table>

where $\theta_k = \alpha_k / \sum_{j=1}^{k} \alpha_j$, and $f^*$ is the set of categories with positive expenditures.

The budget-share Eq. (6) shows how the budget-allocation model relates to the Engel curves discussed earlier. The term $\theta_{m, k}$ in Eq. (6) captures the consumption priority of household $h$ for category $i$ (i.e., upward/downward shifts in the Engel curves in the same budget), while the other terms ($\delta_{s, k}$ and $\sum_{j=1}^{k} \lambda_{m, j}$) capture the budget effect (changes in allocation as the budget grows/shrinks).

Instead of allowing only for observed heterogeneity (i.e., diversity in underlying preferences not accounted for by the observed predictors) in the taste parameter ($\alpha_{m, k}$) for each category $i$, as in Du and Kamakura (2008), we also attempt to explain the households’ consumption priorities using descriptors of the households’ composition and their socioeconomic classification, captured by the predictor vector ($W_{m, h}$).

$$\alpha_{m, h} = \exp(\gamma_i + \delta_{m, h} W_{m, h} + \lambda_{m, h} Z_{m, h} + \epsilon_{m, h}).$$

Details about the estimation of this model can be found in Du and Kamakura (2008). Once the model parameters are estimated for a sample of households, expenditure predictions can be obtained for any household by simulating the budget allocation process. This simulation can be carried out by applying a simple allocation heuristic implied by the model in Eqs. (4)–(7) for each sample household:

1. Start with zero allocations to all categories ($m_{m, h} = \emptyset$) and a full budget.
2. Calculate all categories' marginal utilities $\frac{\partial u_{m, h}}{\partial y_{m, h}} = \exp(\alpha_{m, h})$.
3. Allocate a dollar to the category with the highest marginal utility ($m_{m, h} = \max \{ \alpha_{m, h} \}$ + $1$).
4. Deduct a dollar from the total budget.
5. Repeat steps 2–4 until the budget is totally depleted.

This simple simulation algorithm produces allocations that maximize the estimated utility function subject to the budget constraints faced by each household, allowing us to perform counterfactual analyses that simulate consumption expenditures for each household under different hypothetical conditions, which we will use later to predict how households respond to changes in their consumption budget and how socioeconomic status affect consumption to better understand the differences in consumption priorities across social strata.

4.2. How Brazilians spent their money in 2003 and 2009

Application of the model described in Eqs. (4)–(7) to our data using our proposed socioeconomic stratification produced the parameter estimates reported in Table 7. As shown by Du and Kamakura (2008), the two factor loadings (x1 and x2) along with the household-level latent factor scores (z1 and z2) capture individual differences in preferences that are not explained by demographics and social class. The $\sigma$ parameter measures the error variance associated with the utility function for each consumption category while $\beta$ shows how the utility function changes in relation to the quantity consumed in each category. The response parameters for each of the predictors in the budget allocation model can be interpreted in a similar way to a regression model, except that they refer to shifts in the households’ utility functions due to socio-demographic differences across households in a multinomial logistic regression (see $\theta^*$ in Eq. (6)) and therefore cannot be directly interpreted because they are relative to all consumption categories. Because of these complexities in the interpretation of the model parameters, we rely instead on two counter-factual simulations (using the allocation heuristic discussed earlier) to understand how budget allocations across competing categories are affected by shifts in the total budget and/or differences in consumption priorities across social strata. In the first simulation, we predict each household’s allocation of the same budget using the median budget across all households in our sample, providing insights into the differences in consumption priorities across social strata after equalizing resources constraints. In the second simulation, we increase the budget for each household by 10% and predict the allocation across all consumption categories, thereby measuring the budget effect for each social stratum and consumption category. The results from these two simulations are reported in Table 8. To simplify the presentation of the data we aggregate the simulation results into three socioeconomic segments: the “Upper”
segment, containing strata 1 and 2, the “Middle” segment, containing strata 3 to 5, and the “Lower” segment, aggregating the bottom 3 strata.

The first three columns of results in Table 8 show how the actual consumption budget would be allocated based on the estimated parameters (from Table 7), defining a baseline for comparison. The next 3 columns show how a standard (median) budget would be allocated by all households in the three segments, thereby isolating consumption priorities from budgetary constraints. Comparing these allocated shares with the baseline, differences in allocations observed across social strata are mostly due to differences in consumption priorities rather than differences in the consumption budget, thereby confirming the notion that consumption priorities vary substantially across social strata, either due to class distinctions in “taste” (Bourdieu, 1984), or to the signaling value of conspicuous consumption (Hirsch, 1976; Kamakura & Du, 2012; Schor, 1999). The last six columns of Table 8 quantify the budget effect by showing how the allocations would change if each household had a budget 10% larger than their actual budget. Again, comparing these results to the baseline confirms that most of the differences across the three segments are due to differences in consumption priorities. They also show that the differences in allocations across the upper and middle strata (1 to 5) are due entirely to consumption priorities. In contrast, share allocations for Professional Services, Private Health Insurance and Home Maintenance are expected to increase by 6–7% with a 10% increase in the consumption budget among consumers in the three lowest social strata (6–8), suggesting that these categories are affected the most by the budget constraints faced by consumers in the lower social strata.

5. Conclusions

This study had two main components. In the first portion we develop a framework for social stratification based on the theoretical concepts of social class and of socioeconomic status. The concept of social class implies the categorization of households into a few social classes based on income and the occupation of the head of the household, while socioeconomic status requires the stratification of households along the theoretical construct of permanent income, inferred through a continuous unidimensional wealth index on the basis of observed social indicators such as current income, education and possession of assets. Our proposed framework for social stratification combines the concepts of social class and SES by stipulating latent classes that are empirically identified on the basis of indicator variables measuring both social standing (e.g., occupation, education, access to public services) and permanent income (e.g., per-capita income and possession of assets). Rather than positioning households along a single, unidimensional continuum and arbitrarily partitioning the population along this continuum, our framework avoids the assumption of continuity and unidimensionality in permanent income by specifying latent classes defined on the basis of intuitively appealing constraints on the class parameters. The most valuable practical feature of our proposed social-stratification framework is the fact that because of its latent class formulation, our social stratification framework is robust to missing data, and therefore respondents to different surveys can be classified into our proposed stratification scheme, even when they answer different subsets of socioeconomic indicators, as we demonstrated in Table 5. This seemingly trivial feature allows the marketing researcher to stratify a market using only a subset of the items listed in Table 1, and to relate data from different sources via the social strata, as long as the two sources contain enough socioeconomic indicators out of the 36 listed in Table 1. In contrast, wealth indices, widely applied in commercial marketing research, can only be applied if every respondent in the survey provided data on all indicators.

Obviously, given the dynamism of modern societies and the dramatic shifts in economic power and relevance underway in emerging economies, a framework for social stratification should not be expected to remain relevant for more than a few years, perhaps no longer than a decade. For this reason, the latent class model will have to be re-calibrated, perhaps with the inclusion of new indicators of social standing and permanent income (e.g., access to the internet). However, if the purpose is to analyze changes in social standing over time, a common framework must be used as we have done here to compare social standing between 2003 and 2009 in Brazil.

The second main component of our study utilized the socioeconomic stratification obtained in the first half of this article to understand consumption differences across social strata. Our parameter estimates reported in Table 7 show that socioeconomic stratification explains differences in consumption priorities that go beyond the differences due to household composition. Furthermore, the results from the two counterfactual simulations reported in Table 8 suggest that between the two possible ways that socioeconomic status may affect consumption (budget effect and shifts in consumption priorities), differences in consumption priorities across strata have the strongest effect on consumption. A comparison of consumption expenditures across categories and social strata (Table 6) leads to two interesting and useful insights. First, while the relative size of the social strata changed dramatically between 2003 and 2009, with noticeable shifts towards the upper strata, the concentration of expenditures across strata (measured by the Gini coefficient) did not change as much. This suggests that consumption within each social stratum did not change as much relative to the shifts in stratum sizes. Second, and most importantly, the concentration of expenditures across socioeconomic strata varies considerably across consumption categories, being much less concentrated in essential categories such as Groceries, Starches, and Urban Transportation, and for categories of interest to packaged-goods manufacturers such as dairy, bakery, beverages, hygiene and personal care, and cleaning products. Our results suggest that in these product categories, there is a substantial and growing (based on what we observed between 2003 and 2009) market for brands targeted towards the middle to lower socioeconomic strata, who are probably more price-sensitive and more interested in basic, functional brands, making them a prime market for “good–enough” products/services.

References

Quantifying nation equity with sales data: A structural approach

Junhong Chu

Department of Marketing, NUS Business School, National University of Singapore, 15 Kent Ridge Drive, Singapore 119245, Singapore

A R T I C L E   I N F O

Article history:
First received in 30, June 2011 and was under review for 5 months
Available online 28 September 2012

Keywords:
Nation equity
Country-of-origin effects
Market power
Personal computer market
Structural modeling
Emerging markets
China

A B S T R A C T

Nation equity refers to a product’s equity that is associated with its country of origin (COO). Generalized COO effects have previously been studied through experiments and surveys and measured by quality perception, product attitude, and purchase intention. We propose a structural equilibrium approach to assessing the additional market power COO offers, monetizing nation equity with product sales data, and computing the price premiums and discounts associated with nation equity. We apply this approach to China’s personal computer market between 1995 and 2008 and find that COO does generate additional market power and affects firm markups. We find nation equity to be pervasive, significant, and multidimensional, and it evolves over time, but there is no simple correspondence between market share and nation equity. Large market shares do not necessarily mean positive nation equity and price premiums, negative nation equity and price discounts can evolve into positive nation equity and price premiums, and vice versa. We discuss the implications of our modeling approach and our findings for emerging markets in general, and for China in particular.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Nation equity refers to a product’s equity or its goodwill associated with its country of origin (COO). Generalized COO effects include performance-based effects and normative effects (Maheswaran & Chen, 2006, 2009). Performance-based effects refer to the effects of performance-related country associations on product evaluations (Bilkey & Nes, 1982; Peterson & Jolibert, 1995; Verlegh & Steenkamp, 1999), while normative effects refer to non-performance-related country perceptions, such as the general foreignness effect (Scholder, 1971), normative ethnocentrism (Shimp & Sharma, 1987), and emotional animosity (Klein, Ettenso, & Morris, 1998; Klein, 2002).

Nation equity has been studied primarily with laboratory and survey data and measured by quality perception, product attitude, and purchase intention. Because product attitude and purchase intention do not perfectly predict purchase behavior (Chandon, Morwitz, & Reinartz, 2005), it is important that COO effects and nation equity be investigated with revealed preferences. We propose a structural approach to assess the additional market power that COO offers, to monetize nation equity with sales data, and to compute the associated price premiums and price discounts. We measure nation equity as the difference in the aggregate profits of a country’s manufacturers when a product has a COO versus when it does not, after controlling for the effects of search attributes, brands, and brand-associated non-search attributes. Because we use product sales data, our measure of nation equity reflects the actual choices of consumers and the interactions among and within manufacturers and retailers (Sudhir, 2001; Chu, Chintagunta, & Vilcassim, 2007; Chu & Chintagunta, 2009; Chintagunta, Chu, & Cebollada, 2012). This approach has been used to measure brand equity with transactional data (Goldfarb, Lu, & Moorthy, 2009).

We apply this structural approach to China’s personal computer (PC) market, which contains products from many countries and regions. Using fourteen years of sales data, we obtain interesting insights about the value and dynamics of the COO effect and nation equity in this market. First, we find that COO generates additional market power and significantly affects firm profits. Second, we find nation equity in China, an increasingly globalized and digitized country, to be pervasive and significant, which is contrary to the argument that the COO effect diminishes with globalization and digitization (Samiee, Shimp, & Sharma, 2005). Third, for the first time in this field, we quantify the monetary value of nation equity and compute price premiums associated with positive nation equity and price discounts associated with negative nation equity. We find no simple correspondence between market share and nation equity. Products with a high (low) market share do not necessarily possess positive (negative) nation equity. Fourth, we study the evolution of nation equity, which is difficult and expensive, if not impossible, to accomplish via surveys and experiments. Fifth, we empirically demonstrate the multidimensionality of nation equity in the same market, with different dimensions dominating in different countries. This paper complements the literature on the COO effect by bringing the monetary dimension of nation equity into the picture and studying its magnitude, pattern and evolution with sales data.

The proposed structural approach to quantifying nation equity can be applied to markets such as the European car market where multiple-country ownership of the same brands is prevalent (Brenkers...
& Verboven, 2006). It can be applied to markets such as steel, telecommunications, pharmaceuticals, oil and gas, computer programming and data processing services, advertising services, and insurance, securities, banking and financial services where cross-border mergers and acquisitions (M&As) are common (Kang & Johansson, 2000; Evenett, 2004; Kiene, Helin, & Eckerdt, 2011; Edelstein, Qian, & Tsang, 2011). The approach can also be applied to markets where COO effects dominate and brand effects can be subdued. In markets where both nation equity and brand equity are important and few cross-border M&As are observed, the approach can help to quantify relative nation equity and brand equity. The minimum data required are the sales and prices of products/services from different countries, which have been collected in many markets by various governments, organizations, and agencies.

For emerging markets, building strong nation equity is important both for sustaining domestic consumption and for boosting overseas demand. With economic development and income growth, more consumers in emerging markets start to value product quality. They often perceive products from developed economies to be of higher quality than domestic brands. A recent article in the wsj.com (2012) reports that because of repeated scandals on product quality and safety and because of their increased income, Mainland Chinese consumers have started to loathe “made in China” and cherish imported goods, even though they have a strong desire to support national brands and economy (McEwen, 2007). Mainland Chinese enterprises’ poor performance in brand-building within China has hampered their pace of overseas expansion, even though they have been vying for global markets and global business leadership. Therefore, firms in emerging markets need to focus more on product quality and brand-building to increase their competitive advantage. Our approach to analyzing and interpreting nation equity is particularly useful for governments and enterprises in emerging markets. They are interested in establishing national brands but often mistake products with a high market share as having a high level of quality and thus adopt a low-price strategy to boost the market share of their national brands. Our analysis indicates that a high market share does not imply a high level of nation equity or price premiums. Therefore, our approach can help these governments and firms to better understand the true value of their national brands so that they can adopt measures to enhance their value.

The rest of the paper proceeds as follows. We set up our research framework in Section 2 and summarize the data in Section 3. We describe the econometric model in Section 4 and the estimation and identification in Section 5. We detail the main findings in Section 6 and discuss the managerial implications in Section 7.

### 2. Research framework

The literature has identified the cognitive, affective, and normative roles of a product’s COO in product evaluation and purchase intention (Obermiller & Spangenberg, 1989; Maheswaran, 1994; Gürhan-Canli & Maheswaran, 2000; Batra, Ramaswamy, Alden, Steenkamp, & Ramachander, 2000). A COO can signal overall product quality, as well as specific quality attributes, such as reliability and durability (Li & Wyer, 1994). COOs can have emotional and symbolic value to consumers, including social status and national pride (Botschen & Hemetsberger, 1998). They can link products to a rich set of product images, with sensory, affective, and ritual connotations (Askegaard & Ger, 1998). COOs can also impose social and personal norms on consumers, such as consumer ethnocentrism (Shimp & Sharma, 1987) and consumer animosity (Klein et al., 1998; Klein, 2002). These prior studies confirm the value of COOs to consumers and their effects on different dimensions of demand, such as a consumer’s willingness to pay and choose certain products, which directly affect firm profits. Thus, when quantifying the value of nation equity, we need to account for the effects of COOs on both the demand side and the supply side.

A product is a combination of many aspects: its COO, brand, search attributes, and non-search attributes, such as emotions, culture, trust, imagery, personality, experience, and credence. Some non-search attributes are associated with a product’s brand and become essential inputs for brand equity (Erden & Swait, 1998; Keller, 2003; Goldfarb et al., 2009), while some non-search attributes are associated with its COO (Verlegh & Steenkamp, 1999; Maheswaran & Chen, 2009). When a product is a product’s COO or its nation equity and the additional market power it generates, we need to control for the effects of these attributes. Therefore, we propose to use a structural approach and counterfactual experiments to quantify nation equity and its market power.

When setting wholesale and retail prices, manufacturers and retailers consider the various attributes of products: their COO, brand, search attributes, and non-search attributes. Consumers use these same attributes when deciding which products to purchase to maximize the utility they obtain from the products. The prices of products and their sales volumes affect the profits that firms receive. The aggregated profits of individual firms form a country’s actual profits. When a product’s COO is removed, the COO-associated non-search attributes also disappear. Manufacturers and retailers then set their prices based

---

*Adapted from Goldfarb et al. (2009).

---

Fig. 1. A framework for quantifying nation equity. Adapted from Goldfarb et al. (2009)
on the remaining attributes, and consumers use the same attributes to decide what to buy to maximize their utility. We refer to the profits earned in this situation as “counterfactual profits”. Nation equity is the difference between actual and counterfactual profits (see Fig. 1 for illustration, which is adapted from Goldfarb et al., 2009).

3. Data

Our dataset is from IDC, a global market intelligence firm. It contains details on quarterly PC sales to households and businesses in China from 1995 through 2008 at the manufacturer-brand-model level, along with product attributes such as CPU type, CPU brand, CPU speed, number of CPU cores, and form factor. For example, we observe how many HP Pavilion p6300z desktop PCs with an Intel Pentium dual-core E2xxx 1.5–1.99 GHz chip were sold to households and how many were sold to businesses in Q2 2008 and the average prices in each segment. IDC PC data have been used in Bayus and Putis (1999), Hui (2004), Chu et al. (2007), Kang, Bayus, and Balasubramanian (2010), and many others and are of high quality.

There are 48 PC manufacturers (e.g., Dell, HP) that sell approximately 250 PC brands (e.g., Optiplex, Pavilion) with more than 1000 models (e.g., HP Pavilion p6300z) in the market, but not all PC brands and models exist throughout the entire period. The PCs are of three form factors, desktop, laptop, and ultraportable. Desktop PCs account for 77.37% of the market share and 68.43% of the revenue. Laptop PCs represent 19.92% of the market share and 27.11% of the revenue. Laptop PCs are treated as U.S. products, even though they are assembled outside the U.S. Our survey in 2008 indicated that this is a reasonable assumption. Across five big cities in China, we found that consumers had very good knowledge of the COOs of corporate PC brands: nearly all respondents knew that Lenovo, Founder and Tongfang are Chinese brands, 94% identified Dell as a U.S. brand, 95% identified Toshiba and Sony as Japanese brands, 92% identified Samsung as a Korean brand, and 94% identified Acer as a Taiwanese brand (more results on the survey are reported below).

3.1. PC sales, revenues and prices

PCs represent a large and growing market in China. During 1995 and 2008, 68.41 million PCs worth $46.83 billion were sold to households, and 75.77 million PCs worth $63.15 billion were sold to businesses (Table 1). PC sales grew by double digits in both segments, and higher growth rates were observed in the household segment:

<table>
<thead>
<tr>
<th>Year</th>
<th>Sales (units)</th>
<th>Revenue (million US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>185,938.12</td>
<td>287.40</td>
</tr>
<tr>
<td>1996</td>
<td>252,697.44</td>
<td>361.68</td>
</tr>
<tr>
<td>1997</td>
<td>447,870.49</td>
<td>514.12</td>
</tr>
<tr>
<td>1998</td>
<td>803,442.39</td>
<td>886.71</td>
</tr>
<tr>
<td>1999</td>
<td>1,626,458.13</td>
<td>1,672.21</td>
</tr>
<tr>
<td>2000</td>
<td>3,042,283.77</td>
<td>2,993.60</td>
</tr>
<tr>
<td>2001</td>
<td>3,900,026.02</td>
<td>3,525.65</td>
</tr>
<tr>
<td>2002</td>
<td>4,277,443.92</td>
<td>3,227.35</td>
</tr>
<tr>
<td>2003</td>
<td>4,716,940.12</td>
<td>3,114.25</td>
</tr>
<tr>
<td>2004</td>
<td>5,478,895.90</td>
<td>3,365.22</td>
</tr>
<tr>
<td>2005</td>
<td>6,525,821.36</td>
<td>3,707.98</td>
</tr>
<tr>
<td>2006</td>
<td>8,243,731.50</td>
<td>4,763.61</td>
</tr>
<tr>
<td>2007</td>
<td>12,906,961.88</td>
<td>8,205.59</td>
</tr>
<tr>
<td>2008</td>
<td>15,962,976.17</td>
<td>10,205.10</td>
</tr>
<tr>
<td>Total</td>
<td>68,411,487.21</td>
<td>46,830.56</td>
</tr>
</tbody>
</table>

% Change over previous year

<table>
<thead>
<tr>
<th>Year</th>
<th>Sales</th>
<th>Revenue</th>
<th>Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>57.42</td>
<td>25.85</td>
<td>−20.06</td>
</tr>
<tr>
<td>1997</td>
<td>53.01</td>
<td>42.15</td>
<td>−7.10</td>
</tr>
<tr>
<td>1998</td>
<td>79.39</td>
<td>72.47</td>
<td>−3.86</td>
</tr>
<tr>
<td>1999</td>
<td>102.44</td>
<td>88.59</td>
<td>−6.84</td>
</tr>
<tr>
<td>2000</td>
<td>87.05</td>
<td>79.03</td>
<td>−4.29</td>
</tr>
<tr>
<td>2001</td>
<td>28.19</td>
<td>17.77</td>
<td>−8.13</td>
</tr>
<tr>
<td>2002</td>
<td>9.68</td>
<td>−8.46</td>
<td>−16.54</td>
</tr>
<tr>
<td>2003</td>
<td>10.27</td>
<td>−3.50</td>
<td>−12.50</td>
</tr>
<tr>
<td>2004</td>
<td>16.15</td>
<td>8.06</td>
<td>−6.97</td>
</tr>
<tr>
<td>2005</td>
<td>19.11</td>
<td>10.19</td>
<td>−7.49</td>
</tr>
<tr>
<td>2006</td>
<td>26.32</td>
<td>28.47</td>
<td>1.70</td>
</tr>
<tr>
<td>2007</td>
<td>56.57</td>
<td>72.26</td>
<td>10.02</td>
</tr>
<tr>
<td>2008</td>
<td>23.68</td>
<td>24.37</td>
<td>3.56</td>
</tr>
<tr>
<td>Average*</td>
<td>40.85</td>
<td>31.00</td>
<td>−6.57</td>
</tr>
</tbody>
</table>

* Average is computed as the geometric mean of the series.
the annual rate for households was 40.85%, while it was 24.80% for businesses. PC revenue grew at 31.60% for households and 16.11% for businesses per year. PC prices have been declining at an annual rate of 7%; an average PC cost $1721 in 1995 and $664 in 2008.

3.2. Country-of-origin structure

Table 2 reports the evolution of the COO structure of PC sales in China and mean PC prices. In both segments, Chinese brands account for the majority of PC sales, followed by U.S., Taiwanese, and Japanese brands. The COO structure, particularly the mix of overseas brands, changed over time. Among households, the share of Chinese brands rose from 92.94% in 1995 to a peak of 97.23% in 2001 before declining to 69.03% in 2008. U.S. brands accounted for 5.97% of PC sales in 1995, but dropped to a low of 1.54% by 2002 and increased to 18.25% in 2008. Taiwanese brands dipped to .19% in 1999 but gradually increased but dropped to a low of 1.54% by 2002 and increased to 18.25% in 2008. Japanese brands oscillated between 1% and 5%.

In the business segment, Chinese brands increased from 61.89% in 1995 to 74.99% in 1998 before declining to 61.72% in 2008. U.S. brands dropped from 35.84% in 1995 to 19.68% in 2000 and then fluctuated between 25% and 30%. Taiwanese brands fell to .99% in 1998 and increased to 7.63% in 2008. Japanese brands oscillated between 1% and 5%.

On average, Japanese PCs were priced the highest, followed by “Others”, and U.S. PCs and Chinese PCs were priced the lowest. In some years, U.S. PCs were priced higher than “Others”, while in some other years, the price order was reversed. Over the years, the prices of PCs of all origins have been declining. U.S. PCs experienced the fastest price decline, followed by Chinese PCs. Taiwanese PCs and “Others” had the slowest price decline.

3.3. Market structure

In Table 3, we report the shares of the top 10 PC manufacturers in China. Lenovo, Founder and Tongfang are the top three Chinese firms; Dell, HP and IBM are the top three U.S. firms; Toshiba and Sony are the top two Japanese firms; and Acer and ASUS are the top two Taiwanese firms. In both segments, the shares of the top manufacturers have been changing, reflecting the intense competition in the PC market. Lenovo has dominated the household segment, followed by Founder and Tongfang. In recent years, HP and Dell rose to the second and third positions. Among business buyers in the mid-1990s, Dell, HP (Compaq) and IBM were the most popular PC brands, and Chinese brands were only secondary. However, Chinese brands, led by Lenovo, quickly gained market recognition and acceptance. By 2000, Lenovo and Tongfang became the top two players, and IBM, Dell and HP took the next three positions. Lately, Lenovo has remained as the top brand and acquired the IBM PC Unit in 2004, while Dell and HP have surpassed Tongfang to become the second and third best-selling brands. In terms of market share, Chinese firms are doing better in the household segment than in the business segment. Among U.S. firms, HP has fared better than Dell in the household segment but worse than Dell in the business segment.

3.4. Product preferences and willingness to pay for PCs of different COOs

To assess whether COO plays a role in consumer’s PC purchase decision, we conducted a survey of 1057 respondents in five big Chinese cities (Beijing, Shanghai, Chengdu, Wuhan, and Zhengzhou) in 2008, covering COO knowledge of corporate PC brands, perceived

Table 2
Country-of-origin structure in China's PC market.

<table>
<thead>
<tr>
<th>Market share (%)</th>
<th>Mean prices (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>China</td>
</tr>
<tr>
<td>China</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>92.94</td>
</tr>
<tr>
<td>1996</td>
<td>94.99</td>
</tr>
<tr>
<td>1997</td>
<td>96.57</td>
</tr>
<tr>
<td>1998</td>
<td>95.72</td>
</tr>
<tr>
<td>1999</td>
<td>95.66</td>
</tr>
<tr>
<td>2000</td>
<td>97.12</td>
</tr>
<tr>
<td>2001</td>
<td>97.23</td>
</tr>
<tr>
<td>2002</td>
<td>96.20</td>
</tr>
<tr>
<td>2003</td>
<td>93.91</td>
</tr>
<tr>
<td>2004</td>
<td>91.26</td>
</tr>
<tr>
<td>2005</td>
<td>86.00</td>
</tr>
<tr>
<td>2006</td>
<td>81.46</td>
</tr>
<tr>
<td>2007</td>
<td>75.05</td>
</tr>
<tr>
<td>2008</td>
<td>69.03</td>
</tr>
<tr>
<td>All</td>
<td>82.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>China</th>
<th>US</th>
<th>Japan</th>
<th>Taiwan</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>61.89</td>
<td>.42</td>
<td>1.85</td>
<td></td>
<td>.00</td>
</tr>
<tr>
<td>1996</td>
<td>65.28</td>
<td>2.19</td>
<td>1.73</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>68.83</td>
<td>2.76</td>
<td>1.86</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>74.99</td>
<td>3.13</td>
<td>.99</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>74.84</td>
<td>2.66</td>
<td>1.49</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>73.79</td>
<td>2.35</td>
<td>3.88</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>72.32</td>
<td>2.46</td>
<td>3.82</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>68.67</td>
<td>3.33</td>
<td>2.97</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>63.45</td>
<td>4.77</td>
<td>3.21</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>60.00</td>
<td>5.01</td>
<td>4.05</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>61.40</td>
<td>4.65</td>
<td>5.36</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>59.88</td>
<td>4.38</td>
<td>7.49</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>61.29</td>
<td>3.22</td>
<td>6.80</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>61.72</td>
<td>3.29</td>
<td>7.63</td>
<td>1.59</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>63.64</td>
<td>3.59</td>
<td>5.50</td>
<td>.97</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>China</th>
<th>US</th>
<th>Japan</th>
<th>Taiwan</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>1517.32</td>
<td>1971.21</td>
<td>2740.47</td>
<td>1470.53</td>
<td>1869.02</td>
</tr>
<tr>
<td>1996</td>
<td>1209.74</td>
<td>1719.42</td>
<td>2306.83</td>
<td>1714.83</td>
<td>1278.05</td>
</tr>
<tr>
<td>1997</td>
<td>1129.93</td>
<td>1673.96</td>
<td>1930.63</td>
<td>1512.35</td>
<td>1534.37</td>
</tr>
<tr>
<td>1998</td>
<td>1085.05</td>
<td>1539.34</td>
<td>2259.64</td>
<td>1209.69</td>
<td>1317.34</td>
</tr>
<tr>
<td>1999</td>
<td>1069.34</td>
<td>1441.95</td>
<td>2265.65</td>
<td>1347.29</td>
<td>1885.61</td>
</tr>
<tr>
<td>2000</td>
<td>965.83</td>
<td>1566.52</td>
<td>2890.29</td>
<td>1231.50</td>
<td>2243.51</td>
</tr>
<tr>
<td>2001</td>
<td>888.83</td>
<td>1598.62</td>
<td>2150.84</td>
<td>1047.50</td>
<td>2433.09</td>
</tr>
<tr>
<td>2002</td>
<td>737.96</td>
<td>1218.71</td>
<td>1854.36</td>
<td>929.90</td>
<td>1896.00</td>
</tr>
<tr>
<td>2003</td>
<td>636.05</td>
<td>968.00</td>
<td>1632.00</td>
<td>793.71</td>
<td>1614.34</td>
</tr>
<tr>
<td>2004</td>
<td>581.72</td>
<td>820.38</td>
<td>1597.23</td>
<td>908.06</td>
<td>1462.12</td>
</tr>
<tr>
<td>2005</td>
<td>526.33</td>
<td>687.16</td>
<td>1388.12</td>
<td>821.32</td>
<td>1363.83</td>
</tr>
<tr>
<td>2006</td>
<td>530.60</td>
<td>681.06</td>
<td>1428.06</td>
<td>781.74</td>
<td>1256.10</td>
</tr>
<tr>
<td>2007</td>
<td>577.02</td>
<td>751.59</td>
<td>1372.05</td>
<td>800.60</td>
<td>1061.19</td>
</tr>
<tr>
<td>2008</td>
<td>589.84</td>
<td>726.04</td>
<td>1091.69</td>
<td>686.81</td>
<td>844.32</td>
</tr>
<tr>
<td>All</td>
<td>656.19</td>
<td>770.57</td>
<td>1318.48</td>
<td>763.06</td>
<td>981.81</td>
</tr>
</tbody>
</table>
product quality, willingness to pay, statements on COO-related issues, and PC purchasing behavior. The respondents perceive PCs of different COOs to be of very different quality (Fig. 2). They perceive U.S. PCs to be of the highest quality: 35.73% of the respondents think U.S. PCs have very high quality, 49.11% think they have high quality, and only 1.08% think they have low or very low quality. Japanese PCs come as the second in the perceived quality: 24.20% respondents think Japanese PCs are of very high quality, 51.91% think they have high or very high quality. Chinese and Taiwanese PCs have similar perceived quality at the lower end. Although respondents perceive Chinese PCs to be of lower quality than foreign brands, 82.60% of them strongly agree or agree that the quality of Chinese products has been improving. They do not agree that foreign firms have better customer service than Chinese firms.

Table 3
Sales market shares (%) of major PC manufacturers in China.

<table>
<thead>
<tr>
<th></th>
<th>Lenovo</th>
<th>Founder</th>
<th>Tongfang</th>
<th>U.S.</th>
<th>Dell</th>
<th>HP</th>
<th>IBM</th>
<th>Japan</th>
<th>Sony</th>
<th>Toshiba</th>
<th>Acer</th>
<th>Asus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>4.59</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.81</td>
<td>1.30</td>
<td>.00</td>
<td>.14</td>
<td>.87</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>6.97</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>.20</td>
<td>1.90</td>
<td>.00</td>
<td>.05</td>
<td>.63</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>15.08</td>
<td>3.14</td>
<td>.00</td>
<td>.06</td>
<td>.05</td>
<td>1.61</td>
<td>1.01</td>
<td>.00</td>
<td>.03</td>
<td>.54</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>18.97</td>
<td>4.09</td>
<td>.00</td>
<td>.04</td>
<td>.07</td>
<td>2.27</td>
<td>.00</td>
<td>.03</td>
<td>.35</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>19.86</td>
<td>4.44</td>
<td>.00</td>
<td>.06</td>
<td>.38</td>
<td>2.31</td>
<td>.00</td>
<td>.02</td>
<td>.18</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>29.95</td>
<td>7.97</td>
<td>1.71</td>
<td>.01</td>
<td>.48</td>
<td>1.45</td>
<td>.00</td>
<td>.13</td>
<td>.20</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>27.81</td>
<td>7.80</td>
<td>4.17</td>
<td>.24</td>
<td>.12</td>
<td>.78</td>
<td>.00</td>
<td>.14</td>
<td>.92</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>26.98</td>
<td>7.28</td>
<td>3.85</td>
<td>.63</td>
<td>.15</td>
<td>.65</td>
<td>.05</td>
<td>.38</td>
<td>1.70</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>28.68</td>
<td>9.07</td>
<td>6.03</td>
<td>.45</td>
<td>1.94</td>
<td>.80</td>
<td>.21</td>
<td>.57</td>
<td>1.68</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>27.53</td>
<td>9.99</td>
<td>7.46</td>
<td>.50</td>
<td>4.30</td>
<td>.98</td>
<td>.36</td>
<td>.48</td>
<td>1.05</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>31.99</td>
<td>11.20</td>
<td>7.17</td>
<td>1.10</td>
<td>7.30</td>
<td>.19</td>
<td>1.21</td>
<td>.28</td>
<td>2.02</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>30.83</td>
<td>10.95</td>
<td>7.10</td>
<td>.24</td>
<td>8.03</td>
<td>.00</td>
<td>1.05</td>
<td>.27</td>
<td>3.33</td>
<td>1.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>26.44</td>
<td>7.85</td>
<td>6.91</td>
<td>2.63</td>
<td>13.00</td>
<td>.00</td>
<td>.92</td>
<td>.56</td>
<td>3.56</td>
<td>2.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>24.68</td>
<td>4.81</td>
<td>3.92</td>
<td>5.57</td>
<td>12.50</td>
<td>.00</td>
<td>1.16</td>
<td>1.15</td>
<td>4.41</td>
<td>3.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Lenovo</th>
<th>Founder</th>
<th>Tongfang</th>
<th>U.S.</th>
<th>Dell</th>
<th>HP</th>
<th>IBM</th>
<th>Japan</th>
<th>Sony</th>
<th>Toshiba</th>
<th>Acer</th>
<th>Asus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Businesses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>4.10</td>
<td>.00</td>
<td>.00</td>
<td>1.06</td>
<td>5.14</td>
<td>5.92</td>
<td>.00</td>
<td>.42</td>
<td>1.55</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>6.63</td>
<td>.00</td>
<td>.00</td>
<td>1.05</td>
<td>5.28</td>
<td>7.19</td>
<td>.00</td>
<td>.74</td>
<td>1.70</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>10.02</td>
<td>2.83</td>
<td>.00</td>
<td>1.19</td>
<td>5.11</td>
<td>7.99</td>
<td>.00</td>
<td>1.47</td>
<td>1.65</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>12.54</td>
<td>3.43</td>
<td>.00</td>
<td>1.04</td>
<td>5.35</td>
<td>7.06</td>
<td>.00</td>
<td>1.80</td>
<td>.87</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>19.70</td>
<td>5.70</td>
<td>.00</td>
<td>3.45</td>
<td>6.15</td>
<td>6.83</td>
<td>.00</td>
<td>1.58</td>
<td>1.44</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>24.37</td>
<td>8.14</td>
<td>1.43</td>
<td>5.36</td>
<td>4.77</td>
<td>6.61</td>
<td>.00</td>
<td>1.69</td>
<td>3.69</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>26.09</td>
<td>8.97</td>
<td>2.83</td>
<td>8.15</td>
<td>4.56</td>
<td>6.49</td>
<td>.00</td>
<td>2.16</td>
<td>3.45</td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>21.85</td>
<td>8.79</td>
<td>3.40</td>
<td>10.64</td>
<td>5.28</td>
<td>8.05</td>
<td>.21</td>
<td>2.67</td>
<td>2.71</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>21.59</td>
<td>9.48</td>
<td>5.28</td>
<td>14.68</td>
<td>4.87</td>
<td>7.48</td>
<td>.89</td>
<td>3.30</td>
<td>2.40</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>19.95</td>
<td>9.37</td>
<td>4.60</td>
<td>15.02</td>
<td>6.18</td>
<td>8.24</td>
<td>1.76</td>
<td>2.48</td>
<td>2.45</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>31.37</td>
<td>8.55</td>
<td>2.69</td>
<td>17.11</td>
<td>8.49</td>
<td>1.65</td>
<td>2.27</td>
<td>1.52</td>
<td>2.14</td>
<td>2.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>34.32</td>
<td>9.80</td>
<td>2.71</td>
<td>16.92</td>
<td>10.34</td>
<td>.04</td>
<td>2.35</td>
<td>1.38</td>
<td>3.17</td>
<td>3.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>28.14</td>
<td>7.56</td>
<td>3.56</td>
<td>14.01</td>
<td>13.29</td>
<td>.03</td>
<td>2.01</td>
<td>.79</td>
<td>2.36</td>
<td>3.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>24.90</td>
<td>5.75</td>
<td>2.32</td>
<td>13.12</td>
<td>12.20</td>
<td>.00</td>
<td>1.88</td>
<td>1.14</td>
<td>3.02</td>
<td>4.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Distribution of respondents by perceived PC quality.
focus more on quality control (p-value = .603) (Table 4), but they do agree that foreign firms focus more on quality control (p-value < .0001) and pay more attention to brand image (p-value = .0001) and corporate reputation (p-value < .071). The respondents show strong ethnocentrism. They agree that Chinese should buy national brands (p-value < .0001), and 83.51% of the respondents said they will buy national brands if they have comparable quality to imports. The respondents agree that higher R&D input (p-value < .0001) and higher educational level in a country will lead to higher product quality (p-value = .0015).

We further asked respondents the price premiums they are willing to pay for identically configured foreign PCs over Chinese ones. We find that the respondents are willing to pay an average of 9.02% price premiums (the median is 8.33%) for U.S. PCs. Though the respondents perceive Japanese PCs to be of higher quality, they do not want to pay price premiums but require price compensations (the mean is 7.04%). Of all respondents, 9.18% expressly indicate that they will not buy Japanese PCs. This is a strong sign of animosity toward Japan. Though respondents perceive Taiwanese PCs to be of similar quality as Chinese PCs, they actually require 8.58% price reductions to buy Taiwanese PCs. Korean PCs have slightly higher perceived quality than Chinese PCs, but respondents still require an average of 5.50% price compensation. This is an indication of Chinese consumers’ ethnocentrism in product purchase.

The survey results indicate that COO does play a role in Chinese consumers’ purchasing decisions. Therefore, we need to account for the COO effect when modeling consumer’s PC purchases.

3.5. Shopping channel by COO

In addition to segment-level PC sales, IDC also provided data on PC sales and prices via different distribution channels (retail stores, internet, and so on). However, IDC distributed the segment and channel data separately (no cross-tabulated data were provided); thus, we do not have information on segment–channel combinations, e.g., for the household segment, we do not know how many PCs were sold via retail stores, and how many were sold via the internet. Thus, sales in each customer segment are total sales, and prices are average prices across channels.

A legitimate concern is whether PCs of different COOs are available in different channels and whether the COO effects actually capture country-specific differences in consumer’s channel accessibility. In the survey, we asked the respondents the brand of the PCs they bought most recently and where they bought them. It appears that the shopping channels were available to PCs of all COOs (Table 5), so the availability of a specific channel to PCs of a specific COO should not be the reason for the estimated COO effects.

The data description and our survey results suggest that China’s PC market is an ideal and interesting setting for studying nation equity. First, the PCs sold in China are produced by manufacturers from China, the U.S., Japan, Taiwan, Korea, and European countries (the latter two are grouped as “Others” due to their low market shares). These countries/regions have varying degrees of economic development, cultural affinity, and economic connections with China. These factors have been found to affect nation equity (Maheswaran & Chen, 2009). Second, the COO structure in the market has been changing, enabling us to investigate the dynamic evolution of nation equity. Third, the economic gap between China and the developed world has narrowed substantially in the last 30 years. If there is a positive correlation between product evaluation and economic development (Scholzer, 1991), we should observe China’s nation equity improving. Fourth, the Chinese government has attempted to build national brands, foster nationalism in consumption, and establish the norm “Love China, Buy National”. This may improve the overall equity of Chinese brands. Fifth, because Chinese consumers are found to have a strong and enduring animosity toward Japan (Klein et al., 1998; Klein, 2002), it is worth testing whether this negative feeling manifests itself in purchase behavior. In summary, both performance-based nation equity and normative nation equity are expected to exist and change in China. It is important to see how these forces balance out in actual purchases. Further, China is the world’s largest emerging economy and one of the largest markets. The Chinese market plays a vital role in many multinational companies’ strategic development. Understanding how Chinese consumers evaluate and purchase products from different countries will offer valuable information for multinational companies selling in China. This may also help their strategies in other similar economies.

### Table 4

Responses to COO statements.a

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mean</th>
<th>SD</th>
<th>t-stats</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The quality of Chinese products has been improving</td>
<td>3.990</td>
<td>.987</td>
<td>32.611</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Foreign firms have better customer service</td>
<td>3.016</td>
<td>1.012</td>
<td>.525</td>
<td>.603</td>
</tr>
<tr>
<td>Foreign firms care more about product quality control</td>
<td>3.336</td>
<td>1.003</td>
<td>10.889</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Foreign firms care more about brand image</td>
<td>3.333</td>
<td>1.034</td>
<td>10.461</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Chinese should buy Chinese products</td>
<td>3.057</td>
<td>1.025</td>
<td>1.819</td>
<td>.071</td>
</tr>
<tr>
<td>Chinese should buy Chinese national productsb</td>
<td>3.579</td>
<td>1.075</td>
<td>17.507</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Buying foreign products is equivalent to being non-patrioticc</td>
<td>2.291</td>
<td>1.038</td>
<td>22.219</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Chinese should not buy products from countries unfriendly to China</td>
<td>3.063</td>
<td>1.142</td>
<td>1.792</td>
<td>.077</td>
</tr>
<tr>
<td>The higher the economic development a country has, the higher its product quality</td>
<td>2.849</td>
<td>1.006</td>
<td>4.887</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>The more R&amp;D input a country has, the higher its product quality</td>
<td>3.216</td>
<td>1.036</td>
<td>6.770</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>The higher the education level a country has, the higher its quality</td>
<td>3.100</td>
<td>1.014</td>
<td>3.215</td>
<td>.002</td>
</tr>
</tbody>
</table>

a The questions are asked on a 5-point Likert scale with 1 being “strongly disagree” and 5 being “strongly agree.” The t-tests are against value 3, the neutral point.

b These two are very different statements. The first statement is more normative, while the second has more negative connotation.

c The calculations for this statement were not performed as per the instructions.

### Table 5

<table>
<thead>
<tr>
<th>% Distribution of PC shopping channels by country of origin.</th>
<th>Department store</th>
<th>Electronic store/PC malls</th>
<th>Online</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>7.19</td>
<td>79.77</td>
<td>2.16</td>
<td>2.88</td>
</tr>
<tr>
<td>U.S.</td>
<td>7.50</td>
<td>79.50</td>
<td>9.50</td>
<td>3.50</td>
</tr>
<tr>
<td>Japan</td>
<td>6.25</td>
<td>87.50</td>
<td>3.13</td>
<td>3.13</td>
</tr>
<tr>
<td>Taiwan</td>
<td>6.52</td>
<td>86.96</td>
<td>4.35</td>
<td>2.17</td>
</tr>
<tr>
<td>Others</td>
<td>7.69</td>
<td>76.92</td>
<td>7.69</td>
<td>7.69</td>
</tr>
</tbody>
</table>

1 We limit our analysis to sales via channel intermediaries as they accounted for more than 92% of PC sales in China during 1995 and 2008.
corresponding to their exclusive dealerships (Chu & Chintagunta, 2009; Villas-Boas, 2007). (4) Retailers maximize the profits of all product lines they carry regardless of manufacturer because they have noexclusive dealerships (Sudhir, 2001; Chu & Chintagunta, 2009). We conducted robustness checks on these assumptions (see Section 6.6 for details).

Under these assumptions, the game between manufacturers and downstream firms unfolds in two stages (Besanko, Dubé, & Gupta, 2003). In Stage I, each manufacturer chooses wholesale prices to maximize the profits of their product lines, taking wholesale prices of rival manufacturers as given and taking into account the reactions of retailers. In Stage II, while taking wholesale prices as given, dealers/ VARs choose retail prices to maximize the profits of product lines of each manufacturer, while retailers choose retail prices that maximize the profits of product lines irrespective of their manufacturer. Next, we derive the demand model. The derivation of retailer and manufacturer pricing equations is similar to Chu and Chintagunta (2009), so we relegate it to Appendix A and only report the final equations in the main text.

### 4.2. Random coefficients logit demand model

Similar to Chu et al. (2007) and Goeree (2008), we assume that customer i chooses a PC brand-model $\omega$ to maximize utility. A PC brand $b$ by manufacturer $v$ of country $c$, together with a form factor $f$ and a CPU brand $k$ of log speed $d_{uw}$ with $n_{uw}$ CPU cores by CPU maker $m$ makes a PC model $\omega$. The indirect utility that customer $i$ derives from purchasing model $\omega$ in quarter $t$ is given by:

$$u_{it} = \alpha_{k} + \alpha_{b} + \alpha_{f} + \alpha_{im} + \alpha_{c} + \lambda_{d_{uw}} - \theta_{i}n_{uw} - \beta_{i}(y_{i} - p_{it}) + \gamma_{i}f_{i}(t) + \xi_{it} + \epsilon_{it}. \tag{1}$$

In this specification, we allow unobserved consumer heterogeneity in their intrinsic preference for country ($\alpha_{c}$), manufacturer or corporate brand ($\alpha_{m}$), form factor ($\alpha_{f}$), and CPU maker ($\alpha_{bm}$), as well as in response to CPU speed ($\lambda_{d_{uw}}$), number of CPU cores ($n_{uw}$), and price ($p_{it}$) but not in preferences for specific PC brand ($\alpha_{bm}$) and brand ($\alpha_{m}$) due to their large numbers. Chu et al. (2007) and Chu

### Table 6

<table>
<thead>
<tr>
<th></th>
<th>Household Estimate</th>
<th>Standard deviation</th>
<th>Business Estimate</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>-46.187</td>
<td>24.220</td>
<td>-47.565</td>
<td>20.932</td>
</tr>
<tr>
<td>Others</td>
<td>-34.016</td>
<td>17.454</td>
<td>-37.243</td>
<td>11.010</td>
</tr>
<tr>
<td>Time trend: China/10</td>
<td>-6.936</td>
<td>2.706</td>
<td>-5.904</td>
<td>2.180</td>
</tr>
<tr>
<td>Time trend: US/10</td>
<td>-5.399</td>
<td>2.632</td>
<td>-6.499</td>
<td>1.384</td>
</tr>
<tr>
<td>Time trend: Japan/10</td>
<td>-6.758</td>
<td>2.605</td>
<td>-6.071</td>
<td>3.641</td>
</tr>
<tr>
<td>Time trend: Taiwan/10</td>
<td>-6.868</td>
<td>3.909</td>
<td>-2.878</td>
<td>1.465</td>
</tr>
<tr>
<td>Time trend: Others/10</td>
<td>-4.897</td>
<td>2.520</td>
<td>-8.069</td>
<td>3.241</td>
</tr>
<tr>
<td>Time trend: China2/100</td>
<td>1.185</td>
<td>.519</td>
<td>1.305</td>
<td>.395</td>
</tr>
<tr>
<td>Time trend: US2/100</td>
<td>.564</td>
<td>.359</td>
<td>.777</td>
<td>.314</td>
</tr>
<tr>
<td>Time trend: Japan2/100</td>
<td>.744</td>
<td>.461</td>
<td>.642</td>
<td>.382</td>
</tr>
<tr>
<td>Time trend: Others2/100</td>
<td>.654</td>
<td>.461</td>
<td>.259</td>
<td>.190</td>
</tr>
<tr>
<td>Lenovo</td>
<td>-8.057</td>
<td>3.339</td>
<td>-11.149</td>
<td>5.102</td>
</tr>
<tr>
<td>Founder</td>
<td>-19.978</td>
<td>8.218</td>
<td>-16.537</td>
<td>12.498</td>
</tr>
<tr>
<td>Tingfang</td>
<td>-16.190</td>
<td>4.156</td>
<td>-18.178</td>
<td>10.017</td>
</tr>
<tr>
<td>Dell</td>
<td>-10.785</td>
<td>3.754</td>
<td>-10.507</td>
<td>5.321</td>
</tr>
<tr>
<td>HP</td>
<td>-23.354</td>
<td>8.741</td>
<td>-20.901</td>
<td>7.976</td>
</tr>
<tr>
<td>IBM</td>
<td>-10.018</td>
<td>6.351</td>
<td>-18.511</td>
<td>5.649</td>
</tr>
<tr>
<td>Toshiba</td>
<td>-12.944</td>
<td>8.450</td>
<td>-27.325</td>
<td>17.185</td>
</tr>
<tr>
<td>Sony</td>
<td>-11.811</td>
<td>4.471</td>
<td>-15.547</td>
<td>7.069</td>
</tr>
<tr>
<td>Acer</td>
<td>-13.138</td>
<td>5.965</td>
<td>-17.953</td>
<td>7.114</td>
</tr>
<tr>
<td>AMD</td>
<td>4.293</td>
<td>2.521</td>
<td>-7.373</td>
<td>3.214</td>
</tr>
<tr>
<td>Intel</td>
<td>9.173</td>
<td>1.829</td>
<td>5.063</td>
<td>2.960</td>
</tr>
<tr>
<td>CPU1</td>
<td>12.126</td>
<td>7.857</td>
<td>16.469</td>
<td>10.176</td>
</tr>
<tr>
<td>CPU2</td>
<td>-3.051</td>
<td>1.508</td>
<td>-1.384</td>
<td>3.634</td>
</tr>
<tr>
<td>CPU3</td>
<td>-3.009</td>
<td>3.145</td>
<td>-6.556</td>
<td>3.699</td>
</tr>
<tr>
<td>CPU4</td>
<td>-7.815</td>
<td>4.151</td>
<td>7.813</td>
<td>6.717</td>
</tr>
<tr>
<td>CPU5</td>
<td>-2.608</td>
<td>1.468</td>
<td>-1.637</td>
<td>.912</td>
</tr>
<tr>
<td>CPU6</td>
<td>-2.884</td>
<td>1.303</td>
<td>-1.116</td>
<td>.114</td>
</tr>
<tr>
<td>CPU7</td>
<td>-7.183</td>
<td>2.402</td>
<td>-8.487</td>
<td>3.166</td>
</tr>
<tr>
<td>CPU8</td>
<td>-9.585</td>
<td>1.916</td>
<td>-8.721</td>
<td>3.466</td>
</tr>
<tr>
<td>CPU9</td>
<td>-8.047</td>
<td>1.488</td>
<td>-4.307</td>
<td>3.196</td>
</tr>
<tr>
<td>CPU10</td>
<td>-8.518</td>
<td>3.162</td>
<td>-2.239</td>
<td>1.148</td>
</tr>
<tr>
<td>CPU11</td>
<td>-9.238</td>
<td>1.899</td>
<td>-2.565</td>
<td>2.410</td>
</tr>
<tr>
<td>CPU12</td>
<td>-3.934</td>
<td>6.680</td>
<td>4.300</td>
<td>3.839</td>
</tr>
<tr>
<td>CPU13</td>
<td>-8.959</td>
<td>1.756</td>
<td>-9.702</td>
<td>4.222</td>
</tr>
<tr>
<td>log(CPU speed)</td>
<td>1.718</td>
<td>.267</td>
<td>4.572</td>
<td>1.888</td>
</tr>
<tr>
<td>CPU core</td>
<td>2.393</td>
<td>2.215</td>
<td>2.003</td>
<td>1.442</td>
</tr>
<tr>
<td>Desktop</td>
<td>-1.047</td>
<td>.461</td>
<td>-5.084</td>
<td>1.324</td>
</tr>
<tr>
<td>Laptop</td>
<td>-0.054</td>
<td>2.243</td>
<td>-1.049</td>
<td>.512</td>
</tr>
<tr>
<td>Price</td>
<td>-5.949</td>
<td>.437</td>
<td>5.732</td>
<td>1.171</td>
</tr>
</tbody>
</table>

* Coefficients for brands and other manufacturers are not reported for space reasons.
and Chintagunta (2009) take the same treatment in similar contexts (PC market and server market) for the same reason. Because in the PC market corporate brands often dominate individual brands in consumer's choices (buying a Dell PC versus an HP PC, instead of buying an Optiplex versus a Pavilion), this treatment may not introduce much bias. Further, for logit models with aggregate data, the main objective of accounting for unobserved consumer heterogeneity is to generate flexible substitution patterns. As PC models share COO, manufacturer, form factor, CPU maker, CPU speed, CPU cores and price to various degrees, our model specification is rich enough to generate flexible substitution patterns, as shown in Table 7. We also conduct robustness checks on this (see Section 6.6).

\( y_{it} \) is income\(^2\) and \( p_{it} \) is retail price. \( t \) indicates the quarter in which the sale occurred, \( c \) is a country dummy, and \( \gamma_c \) captures country-specific trends in the intrinsic preference. \( g(t) \) is a function of quarter \( t \), capturing the evolution patterns of country preferences. We use different specifications of \( g(t) \), nonparametric, linear, quadratic, and polynomial, and report the results of the quadratic specification (see Section 6.6 on robustness tests), \( \xi_{it} \) represents unobserved product attributes and \( \epsilon_{it} \) stands for the idiosyncratic utility of individual customers.

In the utility specification, nation equity is determined by parameters \( \alpha_c \) and \( \gamma_c \). \( \alpha_c \) captures the effect of all time-invariant factors that are common to country c's PCs, and \( \gamma_c \) captures the effect of all time-varying factors that are common to country c's PCs but not included in the model. These factors include country c's perceived overall product quality and customer service, emotions such as perceived friendliness or antagonism toward a country and technological sophistication, among others.

In the PC market, the brand effect is very important and must be controlled when the COO effect is being measured. We include manufacturer and brand dummies to capture the brand effect. This raises the issue of how the brand and COO effects can be distinguished. The key identification comes from the cross-country ownership of many brands due to cross-border M&As (see Section 5.4 on identification for more detail).

We include COO, manufacturer, brand, CPU type, CPU brand, CPU speed, CPU cores, form factor, and price as the search attributes for PCs. These attributes are the most crucial ones for PC purchases and have been used as the primary choice drivers in other studies of PC demand (e.g., Hui, 2004; Chu et al., 2007; Goeree, 2008). Because our model specification controls for brand, price, and other product attributes, the estimated nation equity is the net effect of COO on product purchases. This is consistent with business practices in cross-border M&As in which post-M&A firms often keep the original brand names, even though the national origin of the products has changed. For example, Lenovo continues to use the ThinkPad and ThinkCentre brands even though they are now Chinese and not U.S. products, and Acer continues to use the Gateway brand name even though Gateway is now owned by a Taiwanese firm, not a U.S. one. Further, as shown in behavioral studies (e.g., Steenkamp, 1989), it is important to control for other quality cues when examining nation equity.

We assume the effects of product attributes and CPU speed, CPU cores, and price coefficients to vary across consumers as follows:

\[
\begin{align*}
&\alpha_{C0} - N(\alpha_c, \sigma_{C0}^2), \\
&\alpha_{C1} - N(\alpha_c, \sigma_{C1}^2), \\
&\alpha_{C2} - N(\alpha_c, \sigma_{C2}^2), \\
&\alpha_{C3} - N(\alpha_c, \sigma_{C3}^2), \\
&\theta - N(\theta, \sigma_{\theta}^2), \\
&\beta - N(\beta, \sigma_{\beta}^2).
\end{align*}
\]

The utility of the outside good is:

\[
u_{0it} = \alpha_0 - p_{0it} + \epsilon_{0it}.
\]

Define \( \delta_{it} = \alpha_c - \alpha_m + \alpha_f + \alpha_{cm} + \alpha_{m} + \lambda d_{ast} + \theta n_{ast} + \beta p_{ast} + \gamma_c g(t) + \xi_{ast} \) as the mean utility across customers and \( \mu_{ast} = (\alpha_c - \alpha_m) + (\alpha_f - \alpha_m) + (\alpha_{cm} - \alpha_{m}) + (\lambda - \lambda_{ast}) + (\theta - \theta_{ast}) + (\beta - \beta_{ast}) \) as the customer-specific utility. \( \alpha_0 \) is normalized
### 4.3. Downstream firm pricing

The pricing equations for exclusive and nonexclusive downstream firms are, respectively,

\[ P_i = P_i^w - (\Gamma_i' \Delta_i')^{-1} S_i \]  
\[ (5a) \]

\[ P_i = P_i^w - (\Delta_i')^{-1} S_i \]  
\[ (5b) \]

where \( P_i^w \) and \( S_i \) are, respectively, the vector of retail price, wholesale price, and market share; \( \Gamma_i \) is the (symmetric) manufacturer ownership matrix; and \( \Delta_i \) is the matrix of the 1st derivatives of the shares with respect to retail prices, or the retail response matrix. Let \( \rho_i \) be the proportions of sales by exclusive dealers/VARs, which are observable in the data. The final retail prices are the weighted averages across exclusive and nonexclusive downstream firms as follows:

\[ P_i = P_i^w - \left\{ \rho_i (\Gamma_i' \Delta_i')^{-1} + (1 - \rho_i) (\Delta_i')^{-1} \right\} S_i \]  
\[ (5) \]

### 4.4. Manufacturer's product line pricing

Under the Bertrand–Nash assumption, the manufacturer pricing equations are:

\[ P_i^w = MC_i^w - (\Gamma_i' \Delta_i')^{-1} S_i = MC_i^w - (\Gamma_i' \Delta_i')^{-1} S_i \]  
\[ (6) \]

where \( MC_i^w \) is the manufacturer’s marginal costs, and \( \Delta_i \) is the matrix of retail reaction functions, or the matrix of the 1st derivatives of retail prices with respect to wholesale prices. Substituting Eq. (6) into Eq. (5), we have the final retail pricing equations as:

\[ P_i = MC_i^w - (\Gamma_i' \Delta_i')^{-1} S_i - \left\{ \rho_i (\Gamma_i' \Delta_i')^{-1} + (1 - \rho_i) (\Delta_i')^{-1} \right\} S_i \]  
\[ (7) \]

Retail prices can be decomposed into the sum of manufacturer marginal cost \( MC_i^w \), the wholesale margin \( - (\Gamma_i' \Delta_i')^{-1} S_i \), and the retail margin \( - (\rho_i (\Gamma_i' \Delta_i')^{-1} + (1 - \rho_i) (\Delta_i')^{-1}) S_i \). The wholesale margin and the retail margin can be separately identified even without knowing wholesale prices (Sudhir, 2001; Chu & Chintagunta, 2009). Thus, a manufacturer’s marginal costs are:

\[ MC_i^w = P_i + (\Gamma_i' \Delta_i')^{-1} S_i + \left\{ \rho_i (\Gamma_i' \Delta_i')^{-1} + (1 - \rho_i) (\Delta_i')^{-1} \right\} S_i \]  
\[ (8) \]

The retail reaction functions \( \Delta_i \) is:

\[ \Delta_i = [\rho_i (\Gamma_i' \Delta_i')^{-1} + (1 - \rho_i) (\Delta_i')^{-1}] T_i^{-1} \]  
\[ (9) \]

where \( T_i \) is a square matrix with the \( e^{th} \) column defined as \( [\Delta_i' \epsilon_{o,0} + H_{out}[P_i - P_i^w] + [\Delta_i' \epsilon_{o,1}]] \) and \( H_{out} = \partial \Delta_i' / \partial P_i \).

### 5. Estimation and identification

#### 5.1. Estimation of demand parameters

Following the recent literature, we adopt a two-step sequential estimation approach (Nevo, 2001; Chu & Chintagunta, 2009). We first estimate demand parameters by the Generalized Method of Moments (Berry et al., 1995). We then use the demand parameter estimates to obtain the manufacturers’ marginal costs for the pricing equations and project marginal costs onto cost shifters.

Retail prices \( p_{i\text{out}} \) might be correlated with unobserved product attributes \( \epsilon_{o,0} \), leading to a potential price endogeneity problem. We use an instrumental variables technique to account for such endogeneity. In addition to the exogenous product attributes included in the utility to zero for identification purposes. We assume that \( \epsilon_{o,0} \)'s follow an extreme value distribution, which yields the logit choice probability:

\[ s_{i\text{out}} = \frac{\exp(\theta_{i\text{out}} + H_{i\text{out}})}{1 + \sum_{o' = 1}^{O} \exp(\theta_{i\text{out}} + H_{i\text{out}})} \]  
\[ (3) \]

The unconditional market share of PC model \( o \) in quarter \( t \) is:

\[ s_{i\text{out}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(\theta_{i\text{out}} + H_{i\text{out}})}{1 + \sum_{o' = 1}^{O} \exp(\theta_{i\text{out}} + H_{i\text{out}})} \phi(\mu) d\mu. \]  
\[ (4) \]

In summary, our demand model specification and assumptions are similar to Hui (2004), Chu et al. (2007), and Goeree (2008). All these authors take a static approach to model consumer demand for PCs. The large number of manufacturers in the PC market makes a dynamic approach computationally infeasible. Static logit models have also been used to estimate demand for other consumer durables such as cars (Berry, Levinsohn, & Pakes, 1995; Petrin, 2002; Brenkers & Verbong, 2006) and computer servers (Chu & Chintagunta, 2009).
function, we use the following cost shifters as instruments: (1) prices of PCs of the same configuration in the U.S. market, (2) various functions of CPU speed, and (3) current and lagged producer price indices for PC RAM and CPU (www.bls.gov) and their interactions with country dummies to allow for country-specific effects. These instruments together explain 72.1% and 75.7% of price variations, respectively, for households and businesses.

In the demand model specification, we allow consumers to have the option of not choosing any of the PCs to allow the market to contract or expand. For the household segment, we use the number of households (318–440 million) as the base market and household PC ownerships as the scale factor (i.e., scale factor = 1 if the ownership is 10%). For the business segment, we use the number of business employees (90–200 million) as the base market and assume one PC for every two employees. We checked the sensitivity of our results to different scale factors and found that the results are robust to scale factors.

5.2. Estimation of parameters in the pricing equations

After estimating the demand parameters, we compute manufacturers’ marginal costs by Eq. (8). We then pool the data from household and business segments and project these costs onto cost shifters, including dummies for COO ($I_t$), manufacturer ($I_m$), brand ($I_b$), form factor ($I_f$), CPU maker ($I_{cm}$), CPU brand ($I_{cb}$), log (CPU speed), number of CPU cores, and quarter ($I_q$). This approach is the same as that in Chu and Chintagunta (2009).

$$\ln(MC_{it}^{aw}) = \kappa I_t + \kappa_{tm}I_m + \kappa_{tb}I_b + \kappa_{tf}I_f + \kappa_{tm}I_m + \kappa_{tb}I_b + \theta h(I_t)$$

(10)

where $h(I_t)$ is a function of quarter $t$, capturing the trend in marginal costs. We use different specifications, nonparametric, linear, quadratic, and polynomial, and report the nonparametric results.

5.3. Quantifying nation equity and computing price premiums and price discounts

We first compute the nation equity of Country A (e.g., China) over Country B (e.g., the U.S.). We ask the following question, “What would the manufacturers’ profits be if customers’ intrinsic preference for Country A were the same as that for Country B?” We then compute the price premiums resulting from positive nation equity and the price discounts needed to overcome negative nation equity. When the intrinsic preference for Country A is higher than that for Country B, there is positive nation equity for Country A and its brands may acquire price premiums. Conversely, when the intrinsic preference for Country A is lower than that for Country B, there is negative nation equity for Country A and its brands may have to lower prices to attract buyers.

The procedure to quantify nation equity is as follows: (1) set the intrinsic preference and the associated time trend for Country A as well as the marginal costs to those for Country B, (2) compute the prices and shares for the new equilibrium by solving the system of demand and pricing equations, (3) compute the changes in manufacturer profits (actual profits minus counterfactual profits), and (4) compute the corresponding changes in prices. A positive (negative) profit change implies positive (negative) nation equity for Country A, and a positive (negative) price change implies price premiums (discounts).

5.4. Identification

As we estimate the COO effects with brand effects, the key identification for nation equity comes from the cross-country ownerships of many brands in the data, a result of cross-border M&As. In 1996, European PC maker Packard Bell merged with NEC’s PC operations outside Japan and sold PCs to the worldwide market. In 1999, Siemens and Fujitsu formed a joint venture to create a major European computer seller. In 2004, the Chinese firm Lenovo acquired IBM’s PC Unit via a phased process over five years. In 2007, Taiwan-based PC maker Acer acquired U.S. PC maker Gateway. In 2008, Acer acquired Packard Bell. A common practice by these international acquirers is to continue to use the original brand names; thus, the same brands end up being owned by more than one country. For example, before the Lenovo’s acquisition, IBM PCs were products of the U.S. In the two years after the acquisition, some IBM PCs in China’s market were sold as U.S. products and some as Chinese products. From 2007, IBM’s PC brands have been marketed as products of Lenovo (China). The cross-country ownerships of PC brands enable us to separately identify nation equity and brand equity.

Our approach to quantifying nation equity is similar to that used to quantify brand equity in Goldfarb et al. (2009), where the authors include brands (e.g., General Mills) and sub-brands (e.g., Wheaties) in their model. Without within-border M&As, they are able to estimate relative (to the base brand) brand and sub-brand equity. Because we observe cross-border M&As, we are able to estimate nation equity. For brands that are not involved in cross-border M&As, we can estimate their relative brand equity. This will not pose a problem, as our focus in this paper is nation equity.

A related issue is whether the COOs of major PC components, such as the CPU and motherboard and even the operating system, confound the findings. This is very unlikely because there is clear country dominance, even monopoly, among these components. A CPU is arguably the most important PC component, and all CPUs in the dataset are U.S. products. Most operating systems are developed in America, while nearly all motherboards are built in Taiwan. Further, the COOs and even the brands of these components, except for the CPU, are not shown to buyers, so they should not affect the estimation of COO effects.

6. Main findings

In this section, we first discuss the demand parameter estimates and the derived country substitution pattern. We next report cost parameter estimates and the imputed channel, retail, and wholesale margins. We then discuss the COO preference structure and pattern of nation equity. Lastly, we report the magnitude and dynamics of nation equity and the associated price premiums and price discounts.

6.1. Demand parameter estimates

Table 6 shows the major parameter estimates of the mixed logit demand model for the two segments. The upper panel contains the means and standard deviations (heterogeneity distribution) of the intrinsic preferences for COO. The negative sign reflects the small market share of inside goods relative to the outside good, and a larger number indicates stronger initial preference for a country. The middle panel displays the time trends of the intrinsic COO preferences, and the negative linear coefficients and positive quadratic coefficients imply convex time trends, as depicted graphically in Fig. 3a and b and discussed in more detail in Section 5.3. The means and standard deviations of the intrinsic preference for the top 10 manufacturers are reported next in Table 6, with a larger number implying a stronger preference for a manufacturer. The lower panel shows the means and standard deviations of parameter estimates for CPU maker, CPU brand, CPU speed, CPU cores, form factor, and price sensitivity.

Customers have different orders of preference for CPUs. In the business segment, the coefficient for Intel chips is 5.063, for AMD chips is −7.373, and for other chips, which are reserved as the base, is 0. Thus, business buyers prefer Intel chips the most and AMD chips the least. In the household segment, the coefficient for Intel chips is 9.173, 4.293 for AMD chips, and 0 for other chips. Thus, household buyers prefer Intel chips the most, followed by AMD
costs. Consistent with the low-cost image of Chinese products, Chinese PCs have the lowest costs (the smallest coefficient among the five COO coefficients). Taiwanese PCs come next. Japanese PCs have the highest cost (the highest coefficient among the five COO coefficients), followed by the U.S. As expected, desktops have the lowest cost, and ultraportables have the highest cost. Intel-based PCs have higher costs than AMD-based PCs. The positive coefficients for CPU speed and CPU cores mean that the higher the CPU speed a PC has, the higher its marginal cost will be, and the more CPU cores a PC has, the higher its marginal cost will be. As for the time path of cost changes, the coefficient for 1Q1995 is reserved as the base. The increasingly negative coefficients for the quarter dummies indicate that PC costs have experienced exponential declines, and faster declines were observed in the 1990s.

In Table 9, we report the mean, median, and standard deviation of the imputed channel, retail, and wholesale markups for the two customer segments. Note that these numbers cannot be added up because they are computed on different bases. The mean channel markup is 42.43% for households and 43.81% for businesses, the mean retail markup is 21.19% for households and 21.26% for businesses, and the mean wholesale markup is 27.67% and 29.91%, respectively, for the two segments. There are substantial variations across manufacturers in profit margin: Chinese firms generally have 35.75% markups, the highest among all firms, followed by Taiwanese firms, “Others”, and U.S. firms, and Japanese firms have the lowest markups. Chinese firms have high markups primarily because of their low costs (Table 8). The differences in profit margins across firms of different COOs indicate that COO generates additional market power and brings additional profits to firms.

### 6.3. Structure and dynamic evolution of COO preference

We now discuss the top two panels of Table 6 and Fig. 3a and b. Customers in the two segments have different orders of COO preference for PCs. Through the entire period, the U.S. had the largest coefficient, indicating that the U.S. was the most preferred COO for PCs in both segments. This is primarily due to America’s level of economic development and high perceived product quality. In addition, Chinese consumers do not have strong negative feelings toward the U.S. On the other hand, even though Japan had a high level of economic and technological development, it was the least preferred COO for PCs by households throughout and by businesses since 1999. This partly was a reflection of Chinese buyers’ strong animosity toward Japan (Klein et al., 1998; Klein, 2002), which was also confirmed in our survey, where many respondents believed Japanese PCs were of very high quality but expressed a strong reluctance to buy them.

China started low in the preference structure but rose quickly. Although all foreign PCs belong to countries/regions with higher levels of economic development than China, in none of the segments was China the least preferred COO for PCs for the entire period. In the household segment, China had been preferred to Japan throughout the time period and to Taiwan and “Others” since 2006, while in the business segment, China had been preferred to Japan since 1999, to “Others” since 2004, and to Taiwan since 2005. This most likely was due to consumer ethnocentrism. Chinese consumers displayed their nationalism in their purchases, as has been advocated by the Chinese government and various organizations. We also observe a strong foreignness effect in PC purchases among Chinese consumers. “Others” was preferred to China as the COO for PCs in most years, even though the economic development of the major country in the “Others” category, South Korea, and its perceived PC quality, were not far ahead of those of China.

For both segments, intrinsic preferences for all countries, but China was declining and leveling off, implying the gradual commoditization of PCs. PC sales went up over the years. However, they were not driven by customers’ growing intrinsic preference, but by other

---

**Table 9**

Manufacturer and retail markups (%).

<table>
<thead>
<tr>
<th></th>
<th>Households</th>
<th></th>
<th>Businesses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>S.D.</td>
<td>Median</td>
</tr>
<tr>
<td>Channel</td>
<td>39.21</td>
<td>42.43</td>
<td>21.98</td>
<td>37.48</td>
</tr>
<tr>
<td>Retail</td>
<td>19.58</td>
<td>21.19</td>
<td>10.97</td>
<td>18.21</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>23.29</td>
<td>27.67</td>
<td>16.37</td>
<td>21.10</td>
</tr>
<tr>
<td>China</td>
<td>31.43</td>
<td>35.75</td>
<td>21.68</td>
<td>32.74</td>
</tr>
<tr>
<td>Japan</td>
<td>13.89</td>
<td>14.81</td>
<td>10.37</td>
<td>14.01</td>
</tr>
<tr>
<td>Taiwan</td>
<td>23.52</td>
<td>26.46</td>
<td>14.86</td>
<td>24.28</td>
</tr>
<tr>
<td>Others</td>
<td>17.95</td>
<td>20.09</td>
<td>10.92</td>
<td>18.87</td>
</tr>
</tbody>
</table>

Channel markup = (retail price – manufacturer marginal cost)/retail price *100.

Manufacturer markup = (wholesale price – manufacturer marginal cost)/wholesale price *100.

Retail markup = (retail price – wholesale price)/retail price *100.

---

3. Because we include an outside good, i.e., the option of allowing consumers not to choose any of the products in the market, the sum of cross-price elasticities will not be close to the own price elasticity in magnitude. The difference between the two is the shift toward the outside good, i.e., buyers exit the market.
factors, such as fast declining prices and improving product features. The intrinsic COO preferences have changed at different speeds. In both segments, the intrinsic preference for China has grown the fastest, while that for Japan has declined the most. It is hard to believe that the quality or technology of Japan’s PCs declined during this period. A more plausible explanation is China’s animosity against Japan, which was aggravated by the repeated visits to the Yasukuni Shrine by Japanese prime ministers and the revisions of its history textbooks. Our survey confirms the normative role of COO.

Comparing Fig. 3a and b, the evolution of the imputed intrinsic COO preference structure with the observed COO structure (Table 2), we find no simple 1-to-1 correspondence between market share and intrinsic COO preference. A high market share does not necessarily imply high intrinsic preference, and the converse is also true. One striking finding is that, even though Chinese products had the largest market shares in both segments, in neither of them had China been the most preferred COO for PCs. The high market share of Chinese products was primarily driven by other factors such as their low prices, not by customers intrinsically preferring China as the COO for their PCs. This implies that a country could increase its market share by exercising marketing mix tools, such as setting low prices, even though such strategies may not be able to enhance the intrinsic preference for that country’s products.

Because of the differential growth and decline rates in the intrinsic COO preferences, the preference gap between China and the U.S. has narrowed. In the household segment, there was a crossover between COO preferences, the preference gap between China and the U.S. has narrowed. In the household segment, there was a crossover between COO preferences, the preference gap between China and the U.S. has narrowed. In the household segment, there was a crossover between COO preferences.

6.4. Existence of nation equity

We now discuss the monetary dimension of the COO effect, nation equity. Although nation equity has been studied mostly for consumer products, several studies have investigated the phenomenon for industrial buyers (Bilkey & Nes, 1982). In general, institutions are viewed as being more rational and better informed than individual buyers and should be less influenced by extrinsic product cues, such as COO. Although direct comparisons of nation equity effect in industrial and individual buying are scarce, Ahmed and d’Astous (1995) suggest that industrial buyers are less susceptible to COO information. From Table 6, we can see that COO preference and evolution in China’s PC market

<table>
<thead>
<tr>
<th>Table 10</th>
<th>Evolution of China’s nation equity and price premiums/discounts.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nation equity (million U.S.$)^4</td>
</tr>
<tr>
<td></td>
<td>Households</td>
</tr>
<tr>
<td></td>
<td>U.S.  Japan  Taiwan  Others</td>
</tr>
<tr>
<td>1995</td>
<td>−26.66  .74  −1.46  −3.51</td>
</tr>
<tr>
<td>1996</td>
<td>−30.93  1.23  −2.31  −5.93</td>
</tr>
<tr>
<td>1997</td>
<td>−47.43  2.12  −3.44  −9.52</td>
</tr>
<tr>
<td>1998</td>
<td>−70.77  4.26  −5.44  −16.65</td>
</tr>
<tr>
<td>1999</td>
<td>−81.58  10.09  −9.21  −32.36</td>
</tr>
<tr>
<td>2000</td>
<td>−94.93  22.82  −13.15  −57.41</td>
</tr>
<tr>
<td>2001</td>
<td>−113.13  31.98  −10.00  −66.02</td>
</tr>
<tr>
<td>2002</td>
<td>−128.19  41.15  −2.01  −60.54</td>
</tr>
<tr>
<td>2003</td>
<td>−150.65  52.40  −8.75  −50.85</td>
</tr>
<tr>
<td>2004</td>
<td>−159.22  69.94  23.94  −38.59</td>
</tr>
<tr>
<td>2005</td>
<td>−163.27  89.65  44.38  −18.53</td>
</tr>
<tr>
<td>2006</td>
<td>−154.77  114.21  75.96  11.36</td>
</tr>
<tr>
<td>2007</td>
<td>−140.04  176.62  146.64  70.58</td>
</tr>
<tr>
<td>2008</td>
<td>−129.51  215.35  192.51  171.81</td>
</tr>
<tr>
<td>Total</td>
<td>−1491.08  832.56  445.16  −106.15</td>
</tr>
</tbody>
</table>

| Price premiums/discounts (%)^b |         |         |         |         |         |         |         |         |         |         |         |         |
| 1995     | −18.59  2.07  −4.08  −9.80 |         | −20.40  −2.17  −3.76  −10.80 |         |         |         |         |         |         |
| 1996     | −18.98  2.14  −4.01  −10.29 |         | −20.08  −1.92  −4.53  −9.92 |         |         |         |         |         |         |
| 1997     | −19.19  2.33  −3.78  −10.49 |         | −19.63  −1.50  −5.00  −8.90 |         |         |         |         |         |         |
| 1998     | −19.21  2.66  −3.40  −10.39 |         | −19.05  −0.89  −5.18  −7.75 |         |         |         |         |         |         |
| 1999     | −19.04  3.12  −3.85  −10.01 |         | −18.34  −0.17  −5.06  −6.46 |         |         |         |         |         |         |
| 2000     | −18.68  3.71  −3.14  −9.33 |         | −17.50  −0.68  −4.66  −5.05 |         |         |         |         |         |         |
| 2001     | −18.13  4.05  −1.27  −8.36 |         | −16.53  1.67  −3.97  −3.50 |         |         |         |         |         |         |
| 2002     | −17.40  4.83  −0.24  −7.11 |         | −15.42  2.91  −2.98  −1.82 |         |         |         |         |         |         |
| 2003     | −16.48  5.73  −0.76  −5.56 |         | −14.19  4.18  −1.71  −2.1 |         |         |         |         |         |         |
| 2004     | −15.37  6.75  2.31  −3.72 |         | −12.82  5.70  −1.4  −0.79 |         |         |         |         |         |         |
| 2005     | −14.07  7.72  3.82  −1.96 |         | −11.52  6.86  1.72  1.82 |         |         |         |         |         |         |
| 2006     | −12.58  8.96  5.50  −8.2 |         | −9.89  8.16  3.87  2.93 |         |         |         |         |         |         |
| 2007     | −10.90  9.31  7.34  1.53 |         | −8.96  8.70  6.31  3.87 |         |         |         |         |         |         |
| 2008     | −9.04  9.94  8.38  3.48 |         | −7.75  9.18  8.04  4.76 |         |         |         |         |         |         |

^4 Nation equity = actual profits – counterfactual profits.

^b Computed as price changes (%) of actual prices over counterfactual prices. Positive numbers mean price premiums that China can charge over the corresponding country and negative numbers mean price discounts China has to make.
are significant both for households and for businesses, implying that nation equity effect holds both for individual and institutional buyers.

One criticism of studies on nation equity effect is that most studies involve only a single cue, i.e., COO is the only information on which subjects base their evaluations. A single-cue study is bound to yield a significant effect that might not exist in the real world (Peterson & Jolibert, 1995; Verliegh & Steenkamp, 1999). This is less of a concern in our case because our model also includes other product cues such as brand, form factor, CPU maker, brand, color and speed, and price. Our estimated nation equity is the net effect of COO on sales. A second criticism of nation equity studies is that, in many of them, subjects are given only verbal references to products, rather than being shown a tangible product. Schooler (1971) finds that product evaluations may differ depending on whether a tangible product is used. A meta-analysis by Peterson and Jolibert (1995) reveals that the nation equity effect may be artificially inflated in studies that only use verbal product descriptions. Again, this is less of a concern in our case because we use actual sales data, and consumers see the products. In addition to COO information, they also observe many other product attributes.

6.5. Monetary value and evolution of nation equity and the associated price premiums and discounts

The different COO preferences result in different levels of nation equity for each country. Here, we focus on China’s nation equity over other countries. In the top panel of Table 10, we report China’s nation equity over the U.S., Japan, Taiwan and “Others” for both customer segments in each year, computed by the method described in Section 5.3. A positive number means positive nation equity of China over another country. In the household segment, China had a total of negative nation equity over the U.S. and “Others” during 1995 and 2008, but a total of positive nation equity versus Japan and Taiwan (the last row of the top panel). If consumers preferred China as much as the U.S., the aggregate profits of China’s manufacturers from 1995 through 2008 would increase by $1.49 billion over their actual profits in that period. In other words, because of households’ lower intrinsic preference for China over the U.S., the actual total aggregate profits of China’s manufacturers were $1.49 billion less than they would have been. On the other hand, if consumer preferences for China were the same as that for Japan, the aggregate profits of China’s manufacturers would be $833 million less. In other words, because of households’ higher intrinsic preference for China over Japan, the actual total aggregate profits of China’s manufacturers were $833 million more than they would have been. In the business segment, China had a total of $1.51 billion negative nation equity over U.S. during 1995 and 2008, but a total of $670 million positive nation equity over Japan, $203 million over Taiwan, and $261 million over “Others”. Schooler and Wildt (1968) find that consumer bias in COO-related product evaluations is price elastic. For most consumers, country-specific bias simply results in a lowering of the perceived quality of foreign (domestic) products, and compensating price concessions can help re-establish the value comparable to that offered by domestic (foreign) products. Thus, the effect of bias on deciding between domestic and foreign products can be offset by manipulating price differentials, such that prospective buyers will take the value offered by the two to be equally attractive. A McKinsey survey also finds that, of the Chinese who said they have a “strong” or “moderate” preference for foreign brands, nearly half said they would shift to a domestic brand if offered a product of similar quality or price (Dyer, 2007). These findings form the basis for our computation of price premiums and price discounts associated with nation equity.

In the bottom panel of Table 10, we report the price premiums/discounts for China over the U.S., Japan, Taiwan and “Others” for each segment in each year. A positive number indicates the price premium China can charge over a foreign country, while a negative number suggests the price discount China has to make over a foreign country to make its products equally attractive. For example, in 1995, China had to cut its PC prices by 18.59% to make its PCs as attractive to households as U.S. PCs, while China could increase its PC prices by 2.07% and still be as attractive to households as the Japanese PCs. In both segments, China had to incur price discounts of 20%–8% over products from the U.S., but it could charge price premiums of 2%–10% over Japan throughout for households and in most years for businesses.

Studies have found that attitudes toward a country and its products can change over time (Nagashima, 1977). A 2007 McKinsey survey shows that Chinese consumers are increasingly confident about the quality of China’s products: 53% of the 6000 respondents said they prefer Chinese brands, up from 46% when the same survey was conducted in 2005. McKinsey, which conducts similar surveys in a number of countries, commented that this was an unusually large change in sentiment in such a short period of time (Dyer, 2007). Our respondents also agreed that the quality of Chinese products has been greatly improved.

Consistent with the above findings, we find that nation equity and the associated price premiums and discounts in China’s PC market have been changing over time. The declining magnitudes of the negative numbers and the change of sign from negative to positive suggest that China’s position over foreign countries in terms of pricing power had been improving. The price discounts over the U.S. that China had to make declined from 18.59% in 1995 to 9.04% in 2008 for households and from 20.46% to 7.75% for businesses. China’s position over Taiwan and “Others” in pricing power switched during the time course in both segments. China initially had to incur price discounts over Taiwan and “Others”, but with declining magnitudes. In the household segment, China’s price discounts over Taiwan dropped from 4.08% in 1995 to .24% in 2002 and over “Others” decreased from 9.80% to .82% in 2006. In the business segment, China’s price discounts over Taiwan declined from 3.76% to .14% in 2004 and over “Others” dwindled from 10.80% to .21% in 2003. Since then, China started to charge price premiums over these countries/regions. By 2008, China could charge 8.38% price premiums over Taiwan and 3.48% price premiums over “Others” in the household segment and 8.04% and 4.76%, respectively, in the business segment. In the household segment, China could charge increasing price premiums over Japan throughout, from 2.07% in 1995 to 9.94% in 2008. However, in the business market, China first had to incur price discounts and later could charge price premiums over Japan. The magnitudes of price premiums and discounts that China could make in 2008 were in tally with our survey findings.

Nation equity, measured as the differences in actual and counter-factual prices, is the joint effect of observed and new equilibrium price, marginal cost, and market share, so its changes may not follow the same trajectory as price changes. China’s negative nation equity over the U.S. first increased over the years and then decreased, even though the price discounts China had to make had been declining. In the household segment, the negative nation equity increased from $27 million in 1995 to $163 million in 2005, before falling to $130 million in 2008. In the business segment, it increased from $37 million to $160 million in 2003 and then declined to $122 million in 2008. The nation equity over Japan increased from $74 million to $215 million for households and from negative negative $4 million to $258 million for businesses. The nation equity of China over Taiwan and “Others” followed a similar path, starting from negative values and later turning to positive values. The changing magnitudes of price premiums or discounts and nation equity, as well as the switching of signs from negative to positive, reveal the dynamic nature of nation equity, bringing hope to emerging economies. Although at present they may be at a disadvantage in consumer’s intrinsic COO preference and have to incur price discounts to be as competitive as foreign countries, they can improve their position over time.
6.6. Robustness tests

We conduct a series of robustness tests on model assumptions, functional form choices, and estimation methods. In addition to testing various interaction effects, the definition of potential market sizes and the scale factors, and the customer heterogeneity distribution (Chu & Chintagunta, 2009), we also carried out the following tests:

6.6.1. Country of ownership versus country of production

We define COO by country of ownership. The question is whether consumers pay more attention to country of production than to country of ownership. Research has found that nation equity is not affected by multinational production (Verlegh & Steenkamp, 1999). We use Dell to empirically test this. Before 1998, Dell PCs sold in China were assembled outside China. From 1998 on, Dell started to provide the Chinese market with assembled-in-China PCs. We re-estimated our models by coding post-2Q1998 Dell PCs as Chinese products. We found that although this slightly increased the intrinsic preference for China, the change was small in magnitude. More importantly, neither the relative COO preference structure nor the relative time trend changes. We therefore conclude that country-of-production does not qualitatively affect our conclusions.

6.6.2. Time trend of COO preferences

We use different specifications, i.e., nonparametric and polynomial, for the time trend of COO preferences. For the nonparametric specification, we estimate a coefficient for each country and quarter combination with a total of $5 \times 55 = 275$ quarter coefficients (14 years of data, 55 for each country as the coefficient for one quarter is normalized to 0). Only 59 of the quarter coefficients can be precisely identified for the household segment, and only 54 can be identified for the business segment. For the polynomial specification, we use linear, quadratic, and cubic trends and find the quadratic specification yields the best results. We therefore report it in the main text.

6.6.3. Time trend in brand and CPU preferences

We examine the interaction between brand dummies and CPU dummies with linear and quadratic time trends to capture possible changes in preferences for specific brand and CPU, which resulted in more than 400 trend parameters, in addition to other parameters in the original demand model. Even in the simple logit model, less than 5% of the trend parameters could be precisely identified, and some of them were positive and some were negative. Further, the country-specific coefficients for linear and quadratic time trend remain roughly unchanged. Therefore, we believe our measures of COO effect and its trend capture the real COO effect.

6.6.4. Nature of downstream dealerships

We assume dealers/VARs are exclusive while retailers are nonexclusive. We also evaluate two other cases: (i) all downstream firms are exclusive and (ii) all downstream firms are nonexclusive and carry all the product lines of all manufacturers. We find that the estimates of wholesale and retail margins by our current treatment are the most consistent with what the industry experts expected. The estimates of wholesale and retail margins are too low under Case (i) and too high under Case (ii) by industry standards.

6.6.5. Other possible firm conducts

To further corroborate our assumptions on firm behavior, we conduct two robustness checks on the possible firm conducts, one assuming that all firms colluded as one entity, and the other assuming that Chinese firms formed one cartel, foreign firms formed another cartel, and the two cartels were in competition. We find that these assumptions lead to too high firm margins. The mean manufacturer margin was over 65% for the first scenario and approximately 54% for the second scenario.

7. Managerial implications and conclusions

In this paper, we apply the structural modeling approach to monetize nation equity and the associated price premiums and discounts with product sales data. We find that nation equity in China’s PC market to be pervasive and significant, generating additional market power for firms of different COOs. Nation equity is also dynamic, evolving steadily over time. China’s nation equity over other countries/regions has been improving. By the end of the study period, China had positive nation equity and could actually charge price premiums over all countries/regions except for the U.S.

A striking finding is that even though Chinese brands have the largest market shares in household and business segments, in neither of these segments is China the most preferred COO for PCs, indicating that nation equity is not equivalent to market share or sales revenues. So, what is the true value of Chinese products to Chinese consumers? In terms of volume, it seems that Chinese products have won over foreign brands. However, it appeared that Chinese manufacturers achieved this not by making goods with a high intrinsic value to consumers, but by being able to price their goods cheaply because they enjoyed cost advantages over foreign competitors. Thus, a key question is, what will Chinese firms do when they lose their cost advantages? A Gallup poll (McEwen, 2007) found that Chinese consumers preferred to buy Chinese products and that the preference became more pronounced between 2004 and 2006. Chinese consumers showed a strong desire to support national brands and the national economy, but only when “all other things were equal”. They are not fully convinced of the quality of Chinese products. Therefore, product quality and freedom-from-defects are primary considerations for Chinese firms.

Our findings confirm the importance of product quality among Chinese buyers. Low prices can lure consumers but cannot enhance their intrinsic preferences or increase nation equity. More pithily, low prices can help firms win a share of their customers’ wallet, not their hearts. Consumers will walk away when they have concerns over product quality, when prices are no longer cheap, or when they are getting richer, just as the wsj.com (2012) reported. Given the current COO preference structure, China still needs to work hard to build its brand recognition, even among consumers inside China. China is not considered to be the most preferred COO for PCs. This is not a failure of China’s national drive to build national brands. Given its low starting point, China has made tremendous progress toward building national brands. The preference for Chinese products has improved substantively over time. Chinese PCs are preferred to all foreign PCs except for those from the U.S., even though all foreign countries are of higher economic development and more advanced technology.

Now it is time for China to reconsider its branding strategy. Instead of focusing on low prices, China needs to enhance the intrinsic value of its products. This can only be achieved by improving the intrinsic quality of its products. It is clear that Chinese firms cannot simply rely on Chinese consumers’ desire to support local industries. In China’s increasingly competitive market, Chinese firms must continue to address the quality imperative. This finding has implications for manufacturing quality control, customer service, consumer communications, and brand marketing.

China is not alone in building national brands and enhancing national brand equity. Other emerging markets may follow China’s path of development. If China can take various measures to enhance its nation equity, they will be valuable examples for other emerging markets. One important lesson that other emerging markets can learn from China is that a high market share does not necessarily mean high nation equity. Thus, their governments should not simply aim to achieve a high market share; instead, they should target the higher goal of enhancing nation equity. In the beginning, they can use price to enter a market and enlarge their market shares. However, they eventually need high-quality products to develop and maintain a presence in international markets.
Due to limitations in the data, our study has several caveats that can be addressed by future research. First, COO has been used as a signal for unobserved product quality. Because we do not have information on the perceived qualities of PCs made by firms from different countries and their temporal changes, we cannot directly examine how COO preference and nation equity have changed in response to changes in product quality. This is an avenue for future research. It will be interesting to directly investigate the association among COO preference, nation equity, and perceived product quality.

Second, nation equity includes performance-based equity and various dimensions of normative equity. Although we find evidence of the existence of these various dimensions, we are not able to quantitatively decompose nation equity into different components. We hope to conduct further research in this area. Third, because we study nation equity in one market, we are unable to investigate the determinants of nation equity. A promising area of research is to study nation equity in many markets and examine its determinants. Another potential research area is to take a dynamic modeling approach to fully capture the dynamics of nation equity, although given the number of players and products in the PC market, the challenges of such an approach are huge. Further, we examine the nation equity of developed economies, such as China’s nation equity in the U.S. market and its dynamic changes. It might be more valuable for developing economies to understand how their products are received in developed economies.

We hope to further investigate this area.

Due to the large number of brands, we do not account for unobserved consumer heterogeneity in preference for specific brands, which might introduce some bias in parameter estimates. Because we do not have channel-specific prices for each customer segment, we have to use average prices in our modeling. We also do not incorporate the effect of advertising due to data unavailability. We hope to overcome all these limitations in future research.

In summary, although nation equity has been studied extensively via laboratory experiments and field surveys, no study has ever quantified its monetary value with actual sales data. We employ structural modeling to study the existence, pattern, magnitude and dynamics of nation equity and apply the approach to China’s PC market. We take a first step to study nation equity with sales data. We hope similar studies can be conducted for other countries or other markets to generalize the findings.

Acknowledgements

This research is supported by the Singapore Ministry of Education Project grant R-316-000-069-112. We thank the editors and two anonymous reviewers of the International Journal of Research in Marketing for their very constructive comments and suggestions, which helped to improve the quality of the paper tremendously. We are very grateful to Pradeep K. Chintagunta, Durairaj Maheswaran, Yuxin Chen, Xinlei (Jack) Chen, Hongju Liu, Ting Zhu, Perry Chou, and seminar participants at Nanyang Technology University, Hong Kong University of Science and Technology, National University of Singapore, and the 2008 Summer Workshop in Industrial Organization and Management Strategy for valuable comments and suggestions. The usual disclaimer applies.

Appendix A. Derivation of downstream firm and manufacturer pricing equations

A.1. Downstream firm pricing

Similar to Chu and Chintagunta (2009), as we observe the proportions of each PC model’s sales by dealers/VARs ($\hat{\rho}_{\omega t}$) and retailers ($1 - \hat{\rho}_{\omega t}$), we can decompose total sales into two parts: (1) sales by exclusive dealers/VARs—one for each manufacturer and (2) sales by a single non-exclusive retailer, which is assumed to carry all product lines by all manufacturers. The objective function of the exclusive downstream firm $r_0$ is to maximize the profits of product lines by manufacturer $v$:

$$\max_{\{\hat{\rho}_{\omega t}\}} \sum_{\omega=1}^{C} \left( \hat{\rho}_{\omega t} - \hat{\rho}_{\omega t}^{\omega} \right) \hat{\rho}_{\omega t} M_t \quad v = 1, \ldots, V. \quad (A1a)$$

The objective function of the nonexclusive downstream firm $r_0$ is

$$\max_{\{\hat{\rho}_{\omega t}\}} \sum_{\omega=1}^{C} \left( \hat{\rho}_{\omega t} - \hat{\rho}_{\omega t}^{\omega} \right) \hat{\rho}_{\omega t} M_t = \sum_{\omega=1}^{C} \left( \hat{\rho}_{\omega t} - \hat{\rho}_{\omega t}^{\omega} \right) \hat{\rho}_{\omega t} M_t \quad (A1b)$$

where $\hat{\rho}_{\omega t}^{\omega}$ is the wholesale price or retail marginal cost and $M_t$ is the market size. Taking derivatives of Eqs. (A1a) and (A1b) with respect to retail prices, we obtain first-order conditions (FOCs) for all exclusive dealers/VARs (Eq. (A2a)) and the nonexclusive retailer (Eq. (A2b)) in matrix form as:

$$S_t + (\Delta_t^v + \delta^v)^{-1} S_t = 0 \quad (A2a)$$

$$S_t + (\Delta_t^v)^{-1} S_t = 0 \quad (A2b)$$

The corresponding pricing equations are, respectively,

$$P_t = P_t^w - (\Delta_t^v)^{-1} S_t \quad (A3a)$$

$$P_t = P_t^w - (\Delta_t^v)^{-1} S_t \quad (A3b)$$

where $\Delta_t^v$ is the (symmetric) manufacturer ownership matrix with the $(\alpha, \omega')$th element being 1 if both $\omega$ and $\omega'$ belong to the same manufacturer and 0 otherwise. $\Delta_t^v$ is the matrix of the 1st derivatives of the shares with respect to retail prices, or the retail response matrix. Its $(\alpha, \omega')$th element is defined as $\partial S_{\omega'} / \partial P_{\omega t} = \beta_{\omega t} (1 - S_{\omega t})$ if $\omega \neq \omega'$ and $\beta_{\omega t} / \partial P_{\omega t} = \beta_{\omega t} (1 - S_{\omega t})$ if $\omega = \omega'$. The final retail prices are the weighted averages across exclusive and nonexclusive downstream firms as follows:

$$P_t = P_t^w \left( \frac{p_t^w}{p_t} - (\Delta_t^v)^{-1} S_t \right) + (1 - \beta_t) \left( \frac{p_t^w}{p_t} - (\Delta_t^v)^{-1} S_t \right)$$

$$= P_t^w \left( \frac{p_t^w}{p_t} - (\Delta_t^v)^{-1} S_t \right) \quad (A4)$$

A.2. Manufacturer’s product line pricing

Under the Bertrand–Nash assumption, a manufacturer $v$ chooses the wholesale prices for all of its products to maximize product line profits. Its objective function is:

$$\max_{\{w_{\omega t}\}} \sum_{\omega=1}^{C} \left( w_{\omega t} - m_{\omega t} \right) S_{\omega t} M_t - FC_{\text{vt}} \quad (A5)$$

where $m_{\omega t}$ is the manufacturer’s marginal cost and $FC_{\text{vt}}$ are its fixed costs. Taking derivatives of Eq. (A5) with respect to wholesale prices, we obtain manufacturer $v$’s FOCs. Using the manufacturer

\footnote{This assumption can be relaxed if data for individual downstream firms are available.}

\footnote{We have special treatment of M&As. For example, HP was merged with Compaq in 2001. Before the merger, HP and Compaq each maximized its product profits; after the merger, HP and Compaq maximized their joint profits.}
ownership matrix \( \Gamma_t \), we can write FOCs of all manufacturers in matrix form as:

\[
S_t + (\Gamma_t^\top \Delta_t^w)(P_t^w - MC_t^w) = 0
\]  
(A6)

where \( \Delta_t^w \) is the manufacturer response matrix. It is the matrix of the 1st derivatives of shares with respect to wholesale prices, and its \((\omega,\omega')\)th element is defined as \( \partial p_{\omega'}/\partial p_{\omega} \). Changes in wholesale prices first affect retail prices and then market share. Let \( \Delta_t \) be the matrix of retail reaction functions, or the matrix of the 1st derivatives of retail prices with respect to wholesale prices with the \((\omega,\omega')\)th element defined as \( \partial p_{\omega'}/\partial p_{\omega} \). By the chain rule, \( \Delta_t^w = \Delta_t \Delta_t^t \). The manufacturer pricing equations are:

\[
P_t^w = MC_t^w - (\Gamma_t^\top \Delta_t^w)^{-1} S_t = MC_t^w - [(\Delta_t^t)^{-1} S_t].
\]  
(A7)

Substituting Eq. (A7) into Eq. (A4), we obtain the final retail pricing equations as:

\[
P_t = MC_t^w - [(\Delta_t^t)^{-1} S_t - \left\{\rho_t \cdot (\Delta_t^t)^{-1} \right\} + (1 - \rho_t) \cdot (\Delta_t^t)^{-1}] S_t.
\]  
(A8)

Retail prices are the sum of a manufacturer's marginal cost \( MC_t^w \), the wholesale margin \( -[(\Delta_t^t)^{-1} S_t] \) and the retail margin \( -(\rho_t \cdot (\Delta_t^t)^{-1}) + (1 - \rho_t) \cdot (\Delta_t^t)^{-1} \). Thus, a manufacturer's marginal costs are:

\[
MC_t^w = P_t - [(\Delta_t^t)^{-1} S_t + \left\{\rho_t \cdot (\Delta_t^t)^{-1} \right\} + (1 - \rho_t) \cdot (\Delta_t^t)^{-1}] S_t.
\]  
(A9)

A3. Downstream reaction function

To obtain the unknown matrix of retail reaction functions \( \Delta_t \), we fully differentiate the retail FOCs (Eqs. (A3a) and (A3b)) with respect to wholesale prices for the exclusive and nonexclusive downstream firms and then take their weighted averages. The details are as follows:

The \( \omega^{th} \) element of Eq. (A3a) is:

\[
S_{\omega t} + (r_{\omega t} - r_{\omega t}^{\text{ext}}) \left( \frac{\partial S_{\omega t}}{\partial p_{\omega t}} \right)^{-1} \left( \frac{\partial p_{\omega t} - p_{\omega t}^{w}}{\partial p_{\omega t}} \right) = 0.
\]  
(A10)

After taking derivatives of Eq. (A10) with respect to manufacturer \( \nu \)'s wholesale prices, the first term is:

\[
\begin{bmatrix}
\frac{\partial S_{\omega t}}{\partial p_{\omega t}}
\end{bmatrix} + \begin{bmatrix}
\frac{\partial S_{\omega t}}{\partial p_{\omega t}}
\end{bmatrix} \begin{bmatrix}
\frac{\partial S_{\nu t}}{\partial p_{\nu t}}
\end{bmatrix} = 0.
\]  
(A11)

The second term is:

\[
\begin{bmatrix}
\frac{\partial p_{\nu t}}{\partial p_{\nu t}}
\end{bmatrix} - \begin{bmatrix}
\frac{\partial p_{\nu t}}{\partial p_{\nu t}}
\end{bmatrix} = 0.
\]  
(A12)

Define \( H_{\omega t} = \begin{bmatrix}
\frac{\partial S_{\omega t}}{\partial p_{\omega t}}
\end{bmatrix} + \begin{bmatrix}
\frac{\partial S_{\nu t}}{\partial p_{\nu t}}
\end{bmatrix} = 0. \) As the \( \omega^{th} \) column of matrix \( T_t \), putting the set of Eq. (A14) for \( \omega = 1, ..., \Omega_t \) together, we can obtain the following equation for the exclusive downstream firms as:

\[
(\Gamma_t^\top \Delta_t) T_t = \Gamma_t^r \Delta_t^r.
\]  
(A15)

Similarly, for the nonexclusive downstream firm, we have:

\[
\Delta_t T_t = \Delta_t^r.
\]  
(A16)

Thus, the matrix of downstream reaction functions is:

\[
\Delta_t = [\rho_t \cdot (\Gamma_t^\top \Delta_t^t) + (1 - \rho_t) \cdot (\Delta_t^t)^{-1}] T_t^{-1}.
\]  
(A17)

where \( T_t \) is a square matrix with the \( \omega^{th} \) column defined as \( \Delta_{\omega t}^{\text{ext}} + H_{\omega t} \{P_t - P_t^{w} \} + [\Delta_{\omega t}^{\text{ext}} \} \} \) and \( H_{\omega t} = \Delta_t / \partial p_{\omega t} \).

References


Marketing capabilities, institutional development, and the performance of emerging market firms: A multinational study

Jie Wu *

Faculty of Business Administration, University of Macau, Avenida Padre Tomás Pereira, Taipa, Macau, China

**A R T I C L E   I N F O**

Article history:
First received in 23, June 2011 and was under review for 6 months
Available online 5 October 2012

Keywords:
marketing capabilities
performance
institutions
economic growth
legislative systems
individualism
emerging economies

**A B S T R A C T**

Data on 19,653 firms from 73 emerging economies on four continents were analyzed to examine how a firm’s marketing capabilities affect its performance. The results show that the relationship is systematically moderated by the level of institutional development in an emerging market. Economic conditions, legislative institutions and social values all have an impact. Superior marketing capabilities have a stronger performance impact in countries with higher levels of economic development and in individualistic societies. These capabilities have a weaker impact in countries with strong legislative systems.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In recent decades, there has been an unprecedented interest in capabilities and their effect on a firm’s competitive advantage. Capabilities are the accumulated, complex bundles of skills and knowledge embedded in organizational processes (Eisenhardt & Martin, 2000; Helfat & Peteraf, 2003). Previous scholarly research has identified technological capabilities (e.g., Song, Droge, Hanvanich, & Calantone, 2005), operational capabilities (e.g., Worren, Moore, & Cardona, 2002), marketing capabilities (e.g., Kotabe, Srinivasan, & Aulakh, 2002) and management capabilities (e.g., Desarbo, Di Benedetto, Song, & Sinha, 2005) as important. That work has shown empirically that all such capabilities can significantly affect a firm’s performance (e.g., Krasnikov & Jayachandran, 2008).

In spite of the growing consensus that capabilities are critical sources of superior firm performance, the previous research has two important deficiencies. First, most studies have been conducted in developed markets, and only a few were undertaken in emerging markets (Burgess & Steenkamp, 2006). This lacuna is surprising because emerging markets offer a fertile ground for establishing the generalizability of the research findings obtained from developed markets and to assess the extent to which they are specific to the institutional context (Steenkamp, 2005). Emerging markets not only provide a natural laboratory for testing theories and developing new ones, but they also offer practical relevance because success in emerging markets is crucial to the future of many companies (Burgess & Steenkamp, 2006).

The second problem with the body of scholarly work to date has been inattention to the role of institutional environments in shaping the effects of capabilities. Researchers have long recognized that the utility of capabilities is likely to vary with the nature of the market and the social environment (Eisenhardt & Martin, 2000), but previous studies have nevertheless overwhelmingly focused on developed markets where the institutional context can be assumed to vary relatively slightly. This focus represents a serious limitation because institutions in emerging markets normally differ markedly from those typical of developed markets (Burgess & Steenkamp, 2006). Compared with developed markets, emerging markets are characterized by rapid changes in their economic, political and social institutions (Hoskisson, Eden, Lau, & Wright, 2000; Peng, 2003). This volatility renders it less obvious whether firms operating in an emerging market should build market-based capabilities to achieve competitive advantage, considering how fast the institutional environment can change (Kim, Kim & Hoskisson, 2010; Peng, Wang, & Jiang, 2008). It is important, therefore, to look at the hidden assumptions and examine how institutional variations condition the role of firm capabilities.

To address these gaps, this study was designed to link marketing capabilities with firm performance and to examine how the role of marketing capabilities varies among different institutional environments. The study hypothesized that marketing capabilities have a stronger performance impact in more developed countries and in individualistic societies and have a weaker impact in countries with stronger legislative systems. These hypotheses were tested using comprehensive survey data on

* Tel.: +853 8397 8887; fax: +853 2883 8320.
  E-mail address: jiewu@umac.mo.

0167-8116/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijresmar.2012.07.001
19,653 firms from 73 emerging economies. The contribution of this study is threefold. First, this study develops a contingent, institution-based perspective on firm capabilities. This study extends prior academic work to emerging markets and examines to what extent and within what limits capabilities matter in emerging markets. Second, this study contributes to an institution-based view of capabilities by theoretically arguing and empirically showing the moderating effect of economic, legislative, and social institutions on the utility of a firm's capabilities. Third, the findings provide empirical evidence relating capabilities and institutional factors with firm performance in a large number of emerging economies, which generalizes the findings to a broader context.

2. Theoretical development and hypotheses

2.1. Marketing capabilities in emerging markets

Marketing capabilities have long been recognized as one of the key capabilities firms rely on to outperform their competitors and provide superior value to customers (Day, 1994). Compared with technological, operational and other such capabilities, marketing capabilities are less susceptible to imitation and replication due to the tacit and idiosyncratic knowledge involved and its imperfect mobility (Krasnikov & Jayachandran, 2008; Simonin, 1999). Superior capabilities are difficult to observe from the market, difficult to acquire elsewhere and difficult to imitate (Krasnikov & Jayachandran, 2008). These capabilities can thus support a sustainable market advantage (Morgan, Vorhies, & Mason, 2009; Vorhies, Morgan, & Autry, 2008).

Marketing capabilities can be studied in terms of their utility for adaption and integration. The adaption perspective focuses on how marketing capabilities help a firm adapt to the evolving requirements of customers and markets. For example, Day (1994) has emphasized an outside-in process that connects “…organizational processes to the external environment and [enables] the business to compete by anticipating market requirements ahead of competitors [while] creating durable relationships with customers, channel members, and suppliers” (1994: 41). A firm with strong marketing capabilities is better able to target and position its products, identifies customers’ needs better and understands better the factors that influence customer choice (Dutta, Narasimhan, & Rajiv, 1999). The integration perspective on marketing capabilities, conversely, focuses on “the combinative capabilities that derive from the integration of embedded marketing routines and practices” (Vorhies et al., 2009: 1316). Grant (1991) has suggested that reconfiguring and re-integrating internal routines plays an important role in exploiting external opportunities and that both are essential if a firm’s capabilities are to be distinctive. Integrating marketing capabilities will lead to better performance because “…such integration reconfigures competencies, reduces the resources deficiency, and generates new applications” (Song et al., 2005: 262).

A defining feature of emerging markets is the rapid changes in their economic, political, and social institutions (Burgess & Steenkamp, 2006; Hoskisson et al., 2000). A fundamental challenge for firms operating in such environments is to predict the changes and respond to them, which would appear to make the adaption perspective particularly relevant in emerging markets. Consistent with that perspective, the marketing capabilities of the firms examined in this study were defined in terms of their ability to decipher the trajectory of customer needs through effective information acquisition and to respond through marketing planning, investment and execution. To a large extent, this conceptualization reflects a firm’s ability to use its accumulated knowledge regarding the market and customers’ needs to anticipate and respond to events and trends ahead of competitors (Day, 1994). A firm accumulates its market knowledge through learning and experimentation over time. Such market knowledge is distributed across groups and business units within the organization. Therefore, a substantial part of market knowledge, which is tacitly held, is difficult to replicate and supports a market position that is hard to match (Krasnikov & Jayachandran, 2008).

2.2. The role of the institutional environment

The institution-based view of the firm (IBV) has emerged as a useful paradigm for explaining firms’ strategies and competitive advantages in emerging markets (Peng et al., 2008). The IBV suggests that an economy’s institutional environment significantly shapes how firms operate and their performance (Peng et al., 2008; Scott, 1995). As market-supporting institutions develop in emerging economies, firms can rely less on network-based, personal relations-oriented strategies and more on arm’s-length contracts and capability-based strategies (Peng, 2003). The importance of marketing capabilities, therefore, depends on the institutional context in which a firm is operating.

In the same vein, Burgess and Steenkamp have noted that “institutional contexts in emerging markets present significant socioeconomic, demographic, cultural, and regulative departure from the assumptions of theories developed in the Western world and challenge our conventional understanding of constructs and their relations” (2006: 338). Drawing on Scott’s (2001) and North’s (1990) work, Burgess and Steenkamp distinguished three distinct but interrelated institutional subsystems—socioeconomic, cultural, and regulative—each of which is important in emerging markets (see also Etzioni & Lawrence, 1991). Following this stream of research, this study focused on economic, legislative, and social institutions and examined how they moderate the relationship between marketing capabilities and firm performance.

2.2.1. Economic development

Economic development is usually indicated by an economy’s annual GDP per capita (Berry, Guillen, & Zhou, 2010). In economies with low levels of economic development, customers’ purchasing power is usually limited. With limited purchasing power, customers prefer affordable products that offer basic functionality over products with new features at a premium price (Burgess & Steenkamp, 2006; Day & Wensley, 1988). To succeed, a company must minimize its costs of labor, advertising, sales and much else. Marketing capabilities are therefore less influential when the market’s economic development is low.

As the economy progresses, customers’ purchasing power increases, and customers’ preferences diversify. Customers come to prefer better-quality products that address their unique preferences. In such conditions, firms must accurately sense the needs of particular market segments and quickly respond to them. Superior marketing capabilities enable firms to acquire and decipher market information and predict the trajectory of customer preferences better. A firm can accordingly anticipate market requirements ahead of competitors and respond better to customers’ evolving demands (Roth & Jackson, 1995).

**Hypothesis 1.** Marketing capabilities have a stronger effect on firm performance in more developed countries.

2.2.2. Legislative institutions

A market’s “legislative institutions” refers to the structures, processes, and legal rules that regulate the market. The legal system defines the formal structure of rights and obligations in an exchange. An inadequate legal system makes market transactions costly because time and resources (including management attention) must be devoted to gathering information concerning such factors as the financial condition of potential buyers and suppliers, the rationality of competitors, police protection, and security systems, all of which entail substantial costs (Khanna & Palepu, 1997; North, 1990; Wu & Chen, 2012). The quality of a market’s legislative institutions also involves the extent to which legislation and regulations are effectively enforced (Rodriguez, Uhlenbruck, & Eden, 2005). In emerging markets, despite the existence of legal codes, inconsistent and unpredictable legal enforcement can result in the prevalence of unethical or even unlawful behavior (e.g., cheating, false advertising, counterfeiting), which creates high levels of uncertainty in market transactions (Sheng, Zhou, & Li, 2011). Firms operating in an economy with...
weak legislative institutions encounter not only high transaction costs but also high uncertainty.

In such a situation, marketing capabilities help firms to reduce transaction costs and uncertainty in market exchanges. Firms with strong marketing capabilities are better placed to cooperate closely with other channel members, suppliers and customers (Day, 1994; Song et al., 2005). Day has argued that marketing capabilities enable firms to build “durable relationships with customers, channel members, and suppliers” (1994: 41). The durable relationship aligns the incentives of the partners, promotes mutual commitment, and discourages opportunistic behavior (one important source of transaction costs) (Das & Teng, 2000; Wu, 2012). It also helps firms acquire information about the financial situation of customers and suppliers less expensively and thus reduces uncertainty in business exchanges (Uzzi, 1997).

A strong legislative system, by contrast, generates transparency and stability regarding contract enforcement and the boundaries of acceptable behavior (Steenkamp & Geyskens, 2006). Firms can easily acquire information about the credibility of business partners through market intermediaries (Chen & Wu, 2011). The firms can have recourse to a hierarchy of laws and regulations to resolve conflicts with their customers and suppliers (Khanna & Palepu, 1997). The role of marketing capabilities in reducing transaction costs and uncertainty therefore is expected to be less important in countries with a strong legislative system.

Hypothesis 2. Marketing capabilities have a weaker effect on firm performance in countries characterized by stronger legislative systems.

2.2.2. Social institutions

A society’s social institutions are based on the cultural values of the society's members. Though often uncodified, these values guide individual and firm behavior (Hofstede, 2001). This study focused on one important dimension of cultural values—individualism—which has received particular attention in cross-cultural research (Hofstede, 2001; Oyserman, Coon, & Kemmelmeier, 2002; Steenkamp & Geyskens, 2006; Stephan & Uhlman, 2010). Individualism refers to the extent to which people in a society prefer to act as individuals. In individualistic societies, people place their personal goals, motivations, desires and interests ahead of those of others, and the desire for uniqueness and independence is pervasive (Oyserman et al., 2002; Steenkamp & Geyskens, 2006).

People in individualistic societies, for example, value consumption experiences customized to their own unique needs more than people in collectivist societies (Steenkamp & Geyskens, 2006). In such societies, individual benefits and preferences are the priority, and the diversity of those preferences means that the market consists of small segments, each of which is characterized by different needs (Vorhies et al., 2009). In an individualistic society, the challenge for firms is to identify significant segments and develop products to satisfy their particular needs. With accumulated knowledge about a market and its customers, superior marketing capabilities enable a firm to perform this segmentation more accurately and invest more wisely in products that are likely to prove profitable (Steenkamp & Geyskens, 2006).

Hypothesis 3. Marketing capabilities have a stronger effect on firm performance in individualistic societies.

3. Data and methods

3.1. Data

The empirical analyses were based on data collected in a Productivity and Investment Climate Survey conducted by the World Bank. The survey covered 79 countries and 44,000 firms, which provided significant variation in institutional contexts. The World Bank has conducted this survey annually since 2002. Typically, 1,200-1,800 interviews are conducted in larger economies (e.g., Brazil), 360 interviews are conducted in medium-size economies (e.g., Sri Lanka), and for smaller economies (e.g., Mali), 150 interviews are conducted. The sample is selected by stratified random sampling with replacement. Firms are classified as small (5–19 employees), medium (20–99 employees) or large.1 (For more details, see World Bank (2003).)

The survey is administered in on-site, face-to-face interviews with general managers, managing directors, accounting managers, human resource managers, and other relevant company staff. The survey included two sections. The first, answered by the general manager, managing director or owner, focused on the business environment, the investment climate, and business strategies. The second part, answered by an accounting or personnel manager, covered product costs, investment flows, firm performance and workforce statistics.

To overcome common method bias, the information from the second part was used to derive firm performance and the information from the first part to indicate firm capabilities. To reduce the possibility of common method bias further, information regarding economic, legislative, and social institutions was collected from separate sources. Information about each economy’s GDP per capita was collected from International Monetary Fund (IMF) reports for 2002–2006, and legislative system information was collected from Kaufmann, Kraay, and Mastruzzi (2007) for 2002–2006. Individualism indicators were collected from the publications of Hofstede (2001). The Appendix reports the details of the measures and their sources. After dropping observations with missing variables, a sample of 19,653 firms in 73 economies emerged.

3.2. Measures

Firm performance was measured with market share because that is not affected by appropriation problems (Coff, 1999; Coff & Lee, 2003). Appropriation problems occur when a portion of the economic value generated from a firm’s resources and capabilities is not captured by the firm’s owners but is appropriated by other stakeholders, such as top managers. Because performance measures, such as accounting returns and stock prices, are set after stakeholders have had an opportunity to try to extract above-market prices for their contributions, such measures suffer from the appropriation problem and may not reveal the true value generated by the firm’s resources and capabilities. In contrast, market share reflects outcomes before any potential appropriation and thus are not affected by any appropriation problem (Crook, Ketchen, Combs, & Todd, 2008); therefore, market share was used as the measure of firm performance.

Marketing capabilities are normally estimated based on either market research, advertising expenditures (e.g., Dutta et al., 1999) or using scales to quantify the factors underlying such capabilities (e.g., Jayachandran, Hewett, & Kaufman, 2004). However, none of those measures can properly reflect the rapid changes so important in emerging markets, and they say little about how a firm has developed its distinctive competencies in predicting market changes, allocating human resources and making appropriate investments.

Following the lead of Vorhies and Morgan (2005) and Morgan et al. (2009), this study developed a composite measure of marketing capabilities that employed three items: (a) the number of months ahead the firm planned its product mix and target markets; (b) based on those plans, the number of months ahead it allocated the necessary human resources; and (c) the number of months ahead it made the necessary investment. Cronbach’s alpha for these items was 0.83, indicating a high reliability for this construct. Using the information from these indicators, a marketing capabilities indicator was constructed for each firm by averaging the three durations. Previous studies (e.g., Morgan et al., 2009; Vorhies & Morgan, 2005) have shown that this measure is

---

1 In many emerging economies the majority of firms are small, and firms with 100 or more employees are considered engines of job creation. A detailed description of the sample design and sample frame is available at www.enterprisesurveys.org.
highly correlated with other constructs used to capture marketing capabilities and is a valid proxy for marketing capabilities.

Previous work by Berry et al. (2010) has suggested that economic development can be captured adequately by an economy’s GDP per capita. The information on GDP per capita for each economy used in this study was obtained from data for 2002–2006 published by the International Monetary Fund.

National legislative systems have previously been measured in academic studies using information on rule of law from Kauffman and his colleagues’ governance indicators (e.g., Kauffman et al., 2007; Kaufmann, Kraay, & Zoido-Lobatón, 1999) (see, for example, Steenkamp & Geyskens, 2006). Rule of law measures “the respect of citizens and the state for the institutions that govern their interactions.” (Steenkamp & Geyskens, 2006: 142) Rule of law includes several indicators that measure the extent to which people have confidence in and abide by the rules of society and these indicators together measure the success of a society in developing an environment in which fair, predictable, impartial, and enforceable rules form the basis of economic and social interactions (Kaufmann et al., 1999). Therefore, this study measured the legislative system of each economy for 2002 through 2006 based on the information on rule of law obtained from the Kauffman et al. (2007).

An index developed by Hofstede is used widely in academic work to measure cultural individualism (e.g., Steenkamp & Geyskens, 2012). The information on cultural individualism for each economy was obtained from Hofstede’s recent index calculation (Hofstede, 2001).

Several other variables that may influence firm performance were also included in the models as controls. Because large firms often have more resources to devote to improving their performance (Chen & Wu, 2011), two dummy variables were created to indicate that large and medium-size firms with small as the reference group. Next, as older firms may be more likely to be trapped in rigidity and competence traps (Sørensen & Stuart, 2000; Wu, 2012), firm age was also included. Prior studies have shown that in China, government ownership enables a firm to gain exclusive distribution channels, subsidies and tax rebates (Wu, 2011). This property is also likely to be true to some extent in other developing economies; therefore, each firm’s percentage of government ownership was included. Conversely, foreign ownership can offer emerging market firms access to advanced knowledge and expertise (Elango & Pattanaik, 2007); therefore, the percentage of foreign ownership was also included.2 Whether a firm was an exporter (1 = yes; 0 = no) was also included because the need to confront international competition gives exporters an incentive to provide high-quality products at lower prices (Aulakh, Kotabe, & Teegen, 2000). Finally, because the sample covered five years, four year dummy variables were created using 2002 as the reference group.

Table 1 provides summary information concerning the countries, the firms sampled, and the economic development and legislative system scores for each of the 73 economies.

3.3. Statistical modeling

The conceptual model of the relationship between institutional development and marketing capabilities involves both firm and country level variables. The firm-level model (Level 1) was of the form

\[ MS_{ij} = \beta_{0j} + \beta_{1j}MC_{ij} + \beta_{2j}GE_{ij} + \beta_{3j}MSZ_{ij} + \beta_{4j}LSZ_{ij} + \beta_{5j}SO_{ij} + \beta_{6j}FO_{ij} + \beta_{7j}EX_{ij} + \beta_{8j}Year2003_{ij} + \beta_{9j}Year2004_{ij} + \beta_{10j}Year2005_{ij} + \beta_{11j}Year2006_{ij} + u_{ij} \]  

(1)

where \( i \) and \( j \) represent firms and countries, respectively; \( MS \) denotes market share; \( MC \) represents marketing capabilities; \( GDP \) is the GDP per capita, \( LAW \) measures the legislative system and \( IND \) is individualism. \( GE \) denotes the firm’s age variable; \( MSZ \) denotes a medium-sized firm and \( LSZ \) a large one; \( SO \) represents state ownership and \( FO \) foreign ownership. \( EX \) represents an exporter. Year2003, Year2004, Year2005 and Year2006 are the year dummies.

Following the lead of previous studies (e.g., Steenkamp & Geyskens, 2006), the firm-level error term \( u_{ij} \) was assumed to be normally distributed with zero mean and variance \( \sigma^2 \). The random effects \( u_{ij} \) (\( q = 0, \ldots, 11 \)) were multivariate and assumed normally distributed over countries (see more discussion in Steenkamp & Geyskens, 2006). In addition, \( u_{ij} \) is the unique effect of country \( j \) on the intercept (\( \beta_{0j} \)) or slope (\( \beta_{1j}, \beta_{11j} \)).

Because firms are nested within countries, applying ordinary least squares linear regression to such multilevel data would lead to biased estimates with unduly small standard errors for the effects. A multilevel mixed-effects linear model, a generalization of standard linear regression for grouped data, can deal with multiple levels of nested groups by enabling the simultaneous estimation of relationships of variables on two (or more) levels considering both fixed effects and random effects (Steenkamp & Geyskens, 2006). Moreover, a multilevel mixed-effects linear model is versatile in specifying the variance-covariance structure of the random-effects equations. This model also allows fitting the model by performing either residual maximum likelihood estimation or maximum likelihood estimation via EM (expectation-maximization) iterations (Rabe-Hesketh & Skrondal, 2008). Therefore, multilevel mixed-effects models with maximum likelihood estimation were used in the analyses. Following the methods of Steenkamp and Geyskens (2006), the Level 1 predictors within countries were group-mean-centered, whereas the Level 2 predictors were grand-mean-centered.

Endogeneity might be present because marketing capabilities may both affect and be affected by firm performance. To address any potential endogeneity, this study followed Laméy’s recommendations (Laméy, Deleersnyder, Steenkamp, & Dekimpe, 2012) and adopted Heckman two-stage modeling. In the first step, the marketing capabilities of firm \( i \) in country \( j \) were regressed against all the other firms from country \( j \). The explanatory variables in this selection model included marketing intensity and all the other predictors. The predicted values from the first-step estimation were then used to construct the inverse Mills ratio \( \lambda_2 \) (also known as the hazard rate). This ratio was calculated by multiplying the predicted values by \(-1\) and calculating the density and distribution values, and using them in the equation: \( \lambda_2 = f(z_2)/[1- F(z_2)] \), where the \( z_2 \) are predicted values from the first-step model. The inverse Mills ratio was then included as a correction in the second-stage regression models to estimate a firm’s performance (Heckman, 1979).

4. Results

Table 2 presents the mean and standard deviation of each variable and the correlation matrix among them. The magnitude of the correlations among the independent variables was low to medium, suggesting that multicollinearity was not a major concern. This finding is confirmed by the variance of inflation (VIF). The VIF values ranged from 1.05 to 4.36, well below the cutoff threshold of 10, which indicates that there were no serious multicollinearity problems in the models (Hair, Anderson, Tatham, & Black, 1998).
Table 1 (continued)

<table>
<thead>
<tr>
<th>Country</th>
<th>Years</th>
<th>No. of firms</th>
<th>Percent (%)</th>
<th>GDP/per capita(US$)</th>
<th>Legal system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uzbekistan</td>
<td>2002</td>
<td>36</td>
<td>0.18</td>
<td>383</td>
<td>−1.41</td>
</tr>
<tr>
<td>Zambia</td>
<td>2002</td>
<td>153</td>
<td>0.78</td>
<td>350</td>
<td>−0.51</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>19,653</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. GDP is expressed in current U.S. dollars per person. Data are derived by first converting GDP in national currency to U.S. dollars and subsequently dividing it by total population.
b. The value of legislative system is obtained from the country ratings provided in the work of Kaufmann et al. (2007).

Table 3 reports the results of the regression models. M1 is the baseline model, which includes only control variables. M2 adds the main effects of marketing capabilities, GDP per capita, the legislative system and individualism. M3 through M5 include the interactions of marketing capabilities with the three institutional factors. M6 is the full model, including all moderator variables and their interaction terms in a single model. The log-likelihood ratios and Wald chi-squares for these models indicate significant explanatory power. The smaller values of Akaikes information criterion (AIC) and the Bayesian information criterion (BIC) in M6 suggest that the relative goodness of fit was significantly improved in the full model.

Hypothesis 1 predicts that marketing capabilities have a stronger impact on firm performance in countries with high levels of economic development. In M3 and M6, the estimated coefficient of the interaction term of marketing capabilities and GDP/capita was consistently positive and significant ($\beta = 1.69, p \leq 0.001; \beta = 1.67, p \leq 0.001$), indicating that economic development positively moderates the relationship between marketing capabilities and firm performance.

Hypothesis 2 predicts that marketing capabilities have a weaker impact on firm performance in countries with strong legislative systems. The results of M4 and M6 show that the estimated coefficient of the interaction term between marketing capabilities and the legislative system was negative and significant ($\beta = -0.18, p \leq 0.001; \beta = -0.12, p \leq 0.01$), indicating that the quality of an economy’s legislative system negatively moderates the relationship between a firm’s marketing capabilities and its performance. The interaction of marketing capabilities and the legislative system is plotted in Fig. 2. As the figure shows, marketing capabilities had a stronger relationship with firm performance at low levels of the legislative system variable, and this positive relationship becomes weaker with higher levels of the legislative system indicator. Thus Hypothesis 2 was supported.

Hypothesis 3 predicts that marketing capabilities have a stronger impact on firm performance in individualistic societies. M5 and M6 show that the estimated coefficient of the interaction term between marketing capabilities and cultural individualism was positive and significant ($\beta = 0.12, p \leq 0.001; \beta = 0.13, p \leq 0.001$), indicating that individualism positively moderates the relationship between marketing capabilities and firm performance. This effect is plotted in Fig. 3, which shows that marketing capabilities had a weak positive relationship with firm performance when individualism is low, but the positive effect becomes stronger when individualism is high, thereby supporting Hypothesis 3.

The sensitivity of these results was tested in several ways. First, the legislative system was measured using an alternative source — the International Country Risk Guides (ICRG) (Chan, Isobe, & Makino, 2008). The information on law and order for each economy was obtained from the International Country Risk Guides (ICRG) for 2002–2006. An economy was rated 1 if it suffered from a very weak legislative system.
and 6 if it enjoyed a very strong one. All the models were re-estimated. The results using the alternative measure were slightly different, but the main patterns remained unchanged (see Table 4).

Second, as discussed above, the models were fitted using maximum likelihood estimation. One important advantage offered by a multilevel mixed-effects linear model is its versatility in specifying a statistical model for fitting using the variance-covariance structure of the random effects. Therefore, all of the models were re-estimated using restricted maximum likelihood estimation. The results were consistent with those reported. In addition, an exchange

### Table 2
Correlation matrix.

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Market share</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Marketing capabilities</td>
<td>0.06*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 GDP/capita (US$)</td>
<td>0.05*</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Regulative system</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Individualism</td>
<td>−0.11*</td>
<td>0.00</td>
<td>0.10*</td>
<td>−0.41*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Firm age</td>
<td>0.14*</td>
<td>0.04*</td>
<td>0.06*</td>
<td>0.04*</td>
<td>0.05*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Medium size</td>
<td>0.07*</td>
<td>−0.05*</td>
<td>0.04*</td>
<td>−0.05*</td>
<td>−0.02*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Large size</td>
<td>0.14*</td>
<td>0.11*</td>
<td>0.07*</td>
<td>0.15*</td>
<td>−0.21*</td>
<td>0.26*</td>
<td>−0.46*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 State ownership</td>
<td>0.09*</td>
<td>0.01</td>
<td>0.00</td>
<td>0.08*</td>
<td>−0.02*</td>
<td>0.24*</td>
<td>−0.04*</td>
<td>0.21*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Foreign ownership</td>
<td>0.03*</td>
<td>0.09*</td>
<td>0.03*</td>
<td>0.06*</td>
<td>0.01</td>
<td>−0.06*</td>
<td>−0.04*</td>
<td>0.17*</td>
<td>−0.09*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>11 Export</td>
<td>0.05*</td>
<td>0.05*</td>
<td>0.18*</td>
<td>0.09*</td>
<td>−0.08*</td>
<td>0.09*</td>
<td>−0.04*</td>
<td>0.34*</td>
<td>−0.02*</td>
<td>0.20*</td>
<td>1.00</td>
</tr>
<tr>
<td>Mean</td>
<td>20.02</td>
<td>0.15</td>
<td>1783.71</td>
<td>−0.19</td>
<td>36.30</td>
<td>22.34</td>
<td>0.33</td>
<td>0.30</td>
<td>5.77</td>
<td>12.89</td>
<td>0.22</td>
</tr>
<tr>
<td>S.D.</td>
<td>28.09</td>
<td>5.53</td>
<td>1752.85</td>
<td>0.77</td>
<td>13.39</td>
<td>18.26</td>
<td>0.47</td>
<td>0.46</td>
<td>22.20</td>
<td>30.82</td>
<td>0.42</td>
</tr>
</tbody>
</table>

*Indicates significance at the $p \leq 0.05$ level of confidence.

### Table 3
Results of hypotheses testing.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>163.59***</td>
<td>165.98***</td>
<td>166.29***</td>
<td>166.05***</td>
<td>166.05***</td>
<td>166.31***</td>
</tr>
<tr>
<td>Firm age</td>
<td>(71.27)</td>
<td>(64.62)</td>
<td>(64.64)</td>
<td>(64.61)</td>
<td>(64.62)</td>
<td>(64.83)</td>
</tr>
<tr>
<td>Medium size</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
</tr>
<tr>
<td>Large size</td>
<td>−73.05***</td>
<td>−73.53***</td>
<td>−73.69***</td>
<td>−73.55***</td>
<td>−73.51***</td>
<td>−73.71***</td>
</tr>
<tr>
<td>State ownership</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
</tr>
<tr>
<td>Foreign ownership</td>
<td>−0.56***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
</tr>
<tr>
<td>Export</td>
<td>−12.59***</td>
<td>−12.58***</td>
<td>−12.66***</td>
<td>−12.57***</td>
<td>−12.56***</td>
<td>−12.65***</td>
</tr>
<tr>
<td>Year2003</td>
<td>−38.37</td>
<td>−38.96</td>
<td>−39.32</td>
<td>−38.95</td>
<td>−38.94</td>
<td>−39.29</td>
</tr>
<tr>
<td>Year2005</td>
<td>−6.01</td>
<td>−6.49</td>
<td>−6.47</td>
<td>−6.49</td>
<td>−6.50</td>
<td>−6.47</td>
</tr>
<tr>
<td>Year2006</td>
<td>1.88</td>
<td>2.18</td>
<td>2.11</td>
<td>2.20</td>
<td>2.17</td>
<td>2.13</td>
</tr>
<tr>
<td>Marketing capabilities</td>
<td>0.32 (0.38)</td>
<td>0.36 (0.36)</td>
<td>0.36 (0.36)</td>
<td>0.36 (0.36)</td>
<td>0.36 (0.37)</td>
<td>0.37 (0.37)</td>
</tr>
<tr>
<td>GDP/capita</td>
<td>0.24***</td>
<td>0.63***</td>
<td>0.22***</td>
<td>0.21***</td>
<td>0.62***</td>
<td></td>
</tr>
<tr>
<td>Regulative system</td>
<td>0.72</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>Individualism</td>
<td>0.25</td>
<td>0.24</td>
<td>0.26</td>
<td>0.25</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td>MC x GDP/capita</td>
<td>−3.77***</td>
<td>−3.74***</td>
<td>−3.77***</td>
<td>−3.78***</td>
<td>−3.74***</td>
<td></td>
</tr>
<tr>
<td>MC x Regulatory system</td>
<td>1.69***</td>
<td>1.69***</td>
<td>1.69***</td>
<td>1.69***</td>
<td>1.69***</td>
<td>1.69***</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−82279.46</td>
<td>−82207.20</td>
<td>−82149.56</td>
<td>−82186.29</td>
<td>−82180.09</td>
<td>−82144.19</td>
</tr>
<tr>
<td>AIC</td>
<td>164588.92</td>
<td>164452.41</td>
<td>164319.13</td>
<td>164412.57</td>
<td>164400.17</td>
<td>164332.37</td>
</tr>
<tr>
<td>BIC</td>
<td>164707.21</td>
<td>164602.24</td>
<td>164496.85</td>
<td>164570.29</td>
<td>164557.89</td>
<td>164505.86</td>
</tr>
<tr>
<td>d.f.</td>
<td>11</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Wald $\chi^2$</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Prob.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: $N=19,653$, z-scores are in parentheses.

*Signifies significance at the $p \leq 0.05$ (*** $p \leq 0.01$; ** $p \leq 0.001$) level of confidence (two-tailed tests).
Marketing capabilities, economic development and firm performance

Fig. 1. Marketing capabilities, economic development and firm performance.

Marketing capabilities, individualistic societies and firm performance

Fig. 3. Marketing capabilities, individualistic societies and firm performance.

from the assumptions of theories developed for Western economies and provide natural laboratories to test those theories’ assumptions and underlying mechanisms and to generalize their findings and identify boundary conditions (Burgess & Steenkamp, 2006). Unlike previous studies that have focused on firms in developed markets (Morgan et al., 2009; Vorhies et al., 2009) or in a limited number of cases in emerging markets (Fahy et al., 2000), this study examined the interplay of capabilities and institutions using a comprehensive sample of 19,653 firms in 73 emerging economies. The results clearly show that marketing capabilities have a positive relationship with firm performance and that the impact is contingent on a market’s economic, legislative, and social conditions. These findings thus generalize previous academic work to a much broader context by showing the importance of firm capabilities in emerging markets and complement the findings of previous studies by pointing to the limitations of firm capabilities.

These results also extend the institution-based view of competition by developing a contingent perspective. In response to calls for examining the moderating role of institutional contexts (Burgess & Steenkamp, 2006), this study has developed a contingent view of capabilities by assessing how economic, legislative, and social conditions help determine their value. The results show that the effect of good marketing becomes stronger when economic development is more advanced. This finding suggests the importance of economic development in enabling market-based capabilities to function effectively. With the development of the economy, customer purchasing power increases, and consumer preferences diverge. Firms can achieve better performance through investing in marketing capabilities. The results also show that the development of the legislative system weakens the effect of marketing capabilities. As laws and regulations become more transparent, and contract enforcement is more predictable, firms are motivated to invest more resources in developing new products and technologies to attract customers and gain market share. The role of marketing capabilities declines. In addition, marketing capabilities have a stronger effect in individualistic societies. Such societies have diversified preferences, and good marketing enables a firm to sense customers’ specific needs and address them by investing more resources and recruiting and training employees to satisfy them. Taken together, the findings enrich the development of a contingent IBV of capabilities, explaining how they interact with the institutional environment to affect firm performance.

5.1. Managerial implications

These findings have important implications for managers. Conventional wisdom suggests that managers in emerging economies should build network-based personal relationships with partners...
and government officials because their support is critical to firm growth in such situations (Peng & Luo, 2000; Wu & Chen, 2012). However, the results consistently show that marketing capabilities have a significant positive relationship with firm performance across many emerging markets. Managers in emerging markets should, of course, try to build marketing capabilities that will help their company achieve better performance. Equally importantly, they need to understand under what conditions marketing capabilities are more or less effective and become skilled in deploying and developing such capabilities. Marketing capabilities become more effective as economic development progresses, but the effect becomes weaker as the legislative system improves. Firms should therefore pay attention to a market’s social orientation when developing and deploying their marketing capabilities.

These findings also offer valuable implications for policy makers. Many emerging markets are deregulating their economies, aiming to foster globally competent firms. The findings suggest that policy makers should encourage firms to develop different capabilities (including marketing, technology, and operations capabilities) because certain capabilities may become more valuable than others as the institutional environment evolves. The recently implemented policies in many emerging markets which aim to not only improve the legislative system but also to promote innovation among indigenous firms indicate that the governments recognize the importance of this issue, though the effectiveness of their policies is yet to be demonstrated in many cases.

This study has several limitations that, in turn, suggest interesting further hypotheses testing.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>163.59**</td>
<td>165.73**</td>
<td>166.04**</td>
<td>166.17**</td>
<td>165.80**</td>
<td>166.22**</td>
</tr>
<tr>
<td></td>
<td>(71.27)</td>
<td>(64.28)</td>
<td>(64.50)</td>
<td>(64.39)</td>
<td>(64.27)</td>
<td>(64.50)</td>
</tr>
<tr>
<td>Firm age</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
<td>−0.16***</td>
</tr>
<tr>
<td></td>
<td>(−21.75)</td>
<td>(−22.03)</td>
<td>(−22.37)</td>
<td>(−22.32)</td>
<td>(−22.16)</td>
<td>(−22.45)</td>
</tr>
<tr>
<td>Medium size</td>
<td>−6.35***</td>
<td>−6.60***</td>
<td>−6.89***</td>
<td>−6.63***</td>
<td>−5.68***</td>
<td>−6.62***</td>
</tr>
<tr>
<td></td>
<td>(−20.64)</td>
<td>(−20.92)</td>
<td>(−20.98)</td>
<td>(−21.09)</td>
<td>(−20.83)</td>
<td>(−21.06)</td>
</tr>
<tr>
<td>Large size</td>
<td>−73.05***</td>
<td>−73.53***</td>
<td>−73.69***</td>
<td>−73.80***</td>
<td>−73.51***</td>
<td>−73.81***</td>
</tr>
<tr>
<td></td>
<td>(−114.28)</td>
<td>(−115.21)</td>
<td>(−115.77)</td>
<td>(−115.84)</td>
<td>(−115.33)</td>
<td>(−115.94)</td>
</tr>
<tr>
<td>State ownership</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
</tr>
<tr>
<td></td>
<td>(9.26)</td>
<td>(9.31)</td>
<td>(9.31)</td>
<td>(9.24)</td>
<td>(9.26)</td>
<td>(9.27)</td>
</tr>
<tr>
<td>Foreign ownership</td>
<td>−0.56***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
<td>−0.57***</td>
</tr>
<tr>
<td></td>
<td>(−101.42)</td>
<td>(−102.31)</td>
<td>(−102.90)</td>
<td>(−102.90)</td>
<td>(−102.54)</td>
<td>(−103.09)</td>
</tr>
<tr>
<td>Export</td>
<td>−12.50***</td>
<td>−12.58***</td>
<td>−12.66***</td>
<td>−12.63***</td>
<td>−12.56***</td>
<td>−12.67***</td>
</tr>
<tr>
<td></td>
<td>(−38.57)</td>
<td>(−38.96)</td>
<td>(−39.31)</td>
<td>(−39.22)</td>
<td>(−38.94)</td>
<td>(−39.35)</td>
</tr>
<tr>
<td>Year2003</td>
<td>−28.76***</td>
<td>−29.22***</td>
<td>−29.17***</td>
<td>−29.17***</td>
<td>−29.23***</td>
<td>−29.15***</td>
</tr>
<tr>
<td></td>
<td>(−8.03)</td>
<td>(−7.66)</td>
<td>(−7.66)</td>
<td>(−7.62)</td>
<td>(−7.65)</td>
<td>(−7.64)</td>
</tr>
<tr>
<td>Year2004</td>
<td>−6.01</td>
<td>−6.82</td>
<td>−6.81</td>
<td>−6.79</td>
<td>−6.83</td>
<td>−6.79</td>
</tr>
<tr>
<td></td>
<td>(−0.92)</td>
<td>(−1.06)</td>
<td>(−1.05)</td>
<td>(−1.05)</td>
<td>(−1.05)</td>
<td>(−1.05)</td>
</tr>
<tr>
<td>Year2005</td>
<td>1.88</td>
<td>2.90</td>
<td>2.84</td>
<td>2.95</td>
<td>2.89</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.50)</td>
<td>(0.49)</td>
<td>(0.50)</td>
<td>(0.49)</td>
<td>(0.49)</td>
</tr>
<tr>
<td>Year2006</td>
<td>−36.93***</td>
<td>−36.22***</td>
<td>−36.30***</td>
<td>−36.31***</td>
<td>−36.20***</td>
<td>−36.44***</td>
</tr>
<tr>
<td></td>
<td>(−9.69)</td>
<td>(−9.48)</td>
<td>(−9.52)</td>
<td>(−9.50)</td>
<td>(−9.47)</td>
<td>(−9.52)</td>
</tr>
<tr>
<td>Inverse Mills Ratio</td>
<td>−110.25**</td>
<td>−110.40**</td>
<td>−110.61**</td>
<td>−110.73**</td>
<td>−110.48**</td>
<td>−110.76**</td>
</tr>
<tr>
<td></td>
<td>(−154.25)</td>
<td>(−154.99)</td>
<td>(−155.68)</td>
<td>(−155.69)</td>
<td>(−155.29)</td>
<td>(−155.90)</td>
</tr>
<tr>
<td>Marketing capabilities</td>
<td>0.24*** (11.86)</td>
<td>0.63*** (15.20)</td>
<td>0.63*** (15.20)</td>
<td>0.21*** (10.17)</td>
<td>0.40*** (5.91)</td>
<td></td>
</tr>
<tr>
<td>GDP/capita</td>
<td>1.97</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.60)</td>
<td>(0.60)</td>
<td>(0.61)</td>
<td>(0.59)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>Regulative system</td>
<td>1.53</td>
<td>1.33</td>
<td>1.38</td>
<td>1.31</td>
<td>1.31</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>(0.81)</td>
<td>(0.82)</td>
<td>(0.85)</td>
<td>(0.81)</td>
<td>(0.85)</td>
<td>(0.85)</td>
</tr>
<tr>
<td></td>
<td>(−2.13)</td>
<td>(−2.12)</td>
<td>(−2.12)</td>
<td>(−2.13)</td>
<td>(−2.12)</td>
<td>(−2.12)</td>
</tr>
<tr>
<td>MC x GDP/capita</td>
<td>1.69***</td>
<td>(10.76)</td>
<td>1.71***</td>
<td>(5.31)</td>
<td>1.17***</td>
<td>(5.31)</td>
</tr>
<tr>
<td>MC x Regulative system</td>
<td>−0.21*** (10.11)</td>
<td>−0.13*** (10.11)</td>
<td>−0.13*** (10.11)</td>
<td>−0.13*** (10.11)</td>
<td>−0.13*** (10.11)</td>
<td></td>
</tr>
<tr>
<td>MC x Individualism</td>
<td>0.12*** (7.37)</td>
<td>0.13*** (7.22)</td>
<td>0.13*** (7.22)</td>
<td>0.13*** (7.22)</td>
<td>0.13*** (7.22)</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−82279.46</td>
<td>−82207.03</td>
<td>−82149.36</td>
<td>−82156.09</td>
<td>−82179.90</td>
<td>−82133.67</td>
</tr>
<tr>
<td>AIC</td>
<td>164588.92</td>
<td>164542.05</td>
<td>164438.72</td>
<td>164532.17</td>
<td>164399.81</td>
<td>164311.35</td>
</tr>
<tr>
<td>BIC</td>
<td>164547.21</td>
<td>164401.89</td>
<td>164406.44</td>
<td>164505.89</td>
<td>164557.53</td>
<td>164484.84</td>
</tr>
<tr>
<td>df</td>
<td>11</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Wald χ²</td>
<td>24703.07</td>
<td>25025.31</td>
<td>25288.14</td>
<td>25258.11</td>
<td>25147.97</td>
<td>25360.28</td>
</tr>
<tr>
<td>Prob. &gt; χ²</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: N = 19,653, z-scores are in parentheses. *Signifies significance at the p < 0.05 (** p < 0.01; *** p < 0.001) level of confidence (two-tailed tests).
fast-growing emerging economies, many companies consider market share a priority for long-term success (Kotler & Gertner, 2002). However, financial performance dimensions, such as return on assets and profitability, are also relevant. Further research should corroborate the findings with financial indicators.

Third, the study focused on marketing capabilities, but technology and operations capabilities are probably also important in emerging economies, as are network-based resources (Chen & Wu, 2011; Wu & Chen, 2012). Further research should expand the model by considering these alternative capabilities, as well as examining how market-based capabilities and network-based resources interact to affect performance.

In summary, the findings of this study indicate that marketing capabilities are positively related with firm performance in emerging economies and that the effect is moderated by aspects of the economic, legislative, and social environment. Further research is needed to explore the interplay of capabilities, institutions and firm performance.

References


Appendix A. Variables: Sources and operationalization

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market share</td>
<td>(Crook et al., 2008; Coff, 1999)</td>
<td>Productivity and Investment Climate (PICS) Survey A firm’s percentage share of the national market</td>
</tr>
<tr>
<td>Marketing capabilities</td>
<td>(Vorhis &amp; Morgan, 2005; Morgan et al., 2009)</td>
<td>Productivity and Investment Climate (PICS) Survey (α = 0.83) Based on its analyses and prediction of market changes, (a) months ahead the firm planned its product mix and target market, (b) months ahead the firm located necessary human resources, (c) months ahead the firm made necessary investment</td>
</tr>
<tr>
<td>GDP/capita</td>
<td>(Berry et al., 2010)</td>
<td>International Monetary Fund (IMF) GDP per capita, current prices(US$)</td>
</tr>
<tr>
<td>Legislative systems</td>
<td>(Steenkamp &amp; Geyskens, 2006)</td>
<td>Kaufmann et al. (2007) The rule-of-law measure reflects the statistical compilation of information obtained from surveys, nongovernmental organizations, commercial risk-rating agencies, and think tanks</td>
</tr>
<tr>
<td>Individualism</td>
<td>(Steenkamp &amp; Geyskens, 2012)</td>
<td>Hofstede</td>
</tr>
<tr>
<td>Firm age</td>
<td>(Sorenson &amp; Stuart, 2000)</td>
<td>Productivity and Investment Climate (PICS) Survey Years since establishment as of 2009</td>
</tr>
<tr>
<td>Medium-sized</td>
<td>(Chen &amp; Wu, 2011)</td>
<td>Productivity and Investment Climate (PICS) Survey Number of employees between 20 and 99</td>
</tr>
<tr>
<td>Large</td>
<td>(Chen &amp; Wu, 2011)</td>
<td>Productivity and Investment Climate (PICS) Survey Number of employees over 100</td>
</tr>
<tr>
<td>State ownership</td>
<td>(Wu, 2011)</td>
<td>Productivity and Investment Climate (PICS) Survey Percent of state ownership</td>
</tr>
<tr>
<td>Foreign ownership</td>
<td>(Elango &amp; Pattanaik, 2007)</td>
<td>Productivity and Investment Climate (PICS) Survey Percent of foreign ownership</td>
</tr>
<tr>
<td>Exporter</td>
<td>(Aulakh et al., 2000)</td>
<td>Productivity and Investment Climate (PICS) Survey A dummy with the value 1 if the firm is identified as export oriented and 0 otherwise</td>
</tr>
</tbody>
</table>
Functional and experiential routes to persuasion: An analysis of advertising in emerging versus developed markets

Lia Zarantonello \(a,^*\), Kamel Jedidi \(b,^1\), Bernd H. Schmitt \(b,^2,^3\)

\(^{a}\) IÉSEG School of Management (LEM-CNRS), Catholic University of Lille, Socle de la Grande-Arche, 1, Parvis de La Défense, 92044 Paris-La Défense cedex, France

\(^{b}\) Columbia Business School, Columbia University, 3022 Broadway, 518 Uris Hall, New York NY 10027, USA

\(^{1}\) Tel.: +1 212 854 3479; fax: +1 212 854 7647.
\(^{2}\) Tel.: +1 212 854 3468.
\(^{3}\) He is currently Nanyang Visiting Professor at Nanyang Technological University (NTU) in Singapore where he directs the Institute on Asian Consumer Insight.

**ARTICLE INFO**

Article history:
First received in 14, October 2011 and was under review for 6 ½ months
Available online 16 October 2012

Keywords:
Emerging markets
Developed markets
Materialist
Postmaterialist
Advertising appeals
Experiential marketing
Functional marketing

**ABSTRACT**

Should advertising be approached differently in emerging than in developed markets? Using data from 256 television commercial tests conducted by a multinational fast-moving consumer goods (FMCGs) company in 23 countries, we consider two routes of persuasion: a functional route, which emphasizes the features and benefits of a product, and an experiential route, which evokes sensations, feelings, and imaginations. Whereas in developed markets the experiential route mostly drives persuasion, the functional route is a relatively more important driver in emerging markets. In addition, we find a differential impact of local/global and traditional/modern. This finding does not hold for individualistic versus collectivistic ad appeals between emerging and developed markets. We discuss implications of our finding for advertising in emerging markets and for the development of a global consumer culture.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The advertising industry in emerging markets (EMs) is of increasing importance. After the global recession that followed the late-2000s financial crisis, global advertising spending has been on the increase again, but this increase largely stems from the emerging economies in the Asia Pacific, Middle East/Africa, and Latin America regions rather than in the developed markets (DMs) in Europe, the U.S., Australia and Japan. According to a 2011 Nielsen’s report (www.nielsen.com), EMs will continue to lead global ad spending for many years to come, with fast-moving consumer goods (FMCGs) representing the category with the highest expected rate of growth.

Prior research has enriched our understanding of how consumers process and respond to advertisements. However, this research has been conducted almost exclusively in high income, industrialized nations (Burgess & Steenkamp, 2006). There may be important differences in ad processing between DMs and EMs, for example, in the way consumers perceive advertising messages and advertising appeals. Consider an FMCGs company that sells a shampoo, razor, or cleaning product. In EMs, contextual factors affecting the brand (e.g., water availability and purity, bathroom facilities in households, as well as the retail and local selling environment) may be quite different from those in DMs. These factors may affect how consumers perceive the advertisements for these brands—for example, the functional benefits communicated in the ads, the sensory and emotional components, or the various image appeals in the ads.

In this research, we empirically investigate whether consumers in EMs process ads differently than consumers in DMs. We focus specifically on the relative effects of functional and experiential routes of ad persuasion. In addition, we investigate the effects of socio-cultural ad appeals on ad processing in EMs and DMs, including perceived referential appeals (local versus global), innovativeness appeals (modern versus traditional), and group-related cultural appeals (individualistic versus collectivistic).

2. Conceptual framework and hypothesis development

2.1. Functional and experiential approaches in advertising

At a broad level, marketing researchers (e.g., Vakratsas & Ambler, 1999) have created an information processing framework of the ad persuasion process in which the advertising message (i.e., the input of the process) generates an internal consumer response, which, in turn, affects consumer behavior (i.e., the output). According to some models
(e.g., Barry & Howard, 1990), advertising results in, and should be measured in, specific behaviors (product purchase, trial, and adoption), while other models suggest measuring ad impact in terms of attitude formation and change (Copper & Croyle, 1984; Olson & Zanna, 1993; Petty & Wegener, 1997; Tesser & Shaffer, 1990).

A large body of research has concentrated on the link between the type of ad message and the internal response. Broadly speaking, an advertising message can be described in terms of its functional-rational or emotional-experiential components (Heath, 2011). The two types of messages have been referred to in various ways in the advertising literature, such as “informational” versus “transformational” (Rossiter & Percy, 1987), “utilitarian” versus “value-expressive” (Johar & Sirgy, 1991), “hard-sell” versus “soft-sell” (Okazaki, Mueller, & Taylor, 2010), and “central” versus “peripheral” messages (Petty & Cacioppo, 1986). In this paper, we will use the terms “functional” and “experiential.” The functional aspects of an ad include the utilitarian references to product features (e.g., attributes, applications, and performance) as well as the benefits and value generated from these features, resulting in a cognitive consumer response (e.g., evaluation) (Abernethy & Franke, 1996). In contrast, the experiential aspects of an ad evoke sensations, feelings, emotions, imaginations, and lifestyles, thus resulting in an affective response (e.g., liking) (Brakus, Schmitt, & Zaranontello, 2009; Holbrook & Hirschman, 1982; Schmitt, 1999).

It should be noted that almost all ads (and certainly the ones used in our empirical studies) include, to some degree, both functional and experiential components. Moreover, the two approaches (targeting cognitions with the functional ad component and targeting affect with the experiential component) may be viewed as two different routes of persuasion (Petty & Cacioppo, 1986). As routes of persuasion, they are not mutually exclusive: advertising communications can adopt either one of the two approaches, or both; in the latter case, cognitive and affective responses are activated simultaneously (De Pelsmacker, Geuens, & Van den Bergh, 2007). Finally, the two internal consumer responses (cognitive and affective) may be related: a positive, cognitive evaluation may, in itself, trigger affect; conversely, an affective response or feeling may trigger a reflective cognitive response to explain its source or justify why the feeling occurred (Chaiken, 1980; Forgas, 1995; Petty & Cacioppo, 1986).

2.2. Ad processing differences across markets

Turning to the central question of this research, do we expect any differences in the effectiveness of functional and experiential routes to persuasion between DMs and EMs? To answer this question, it must be addressed in the context of the broader changes occurring in DMs and EMs.

In his influential work, Inglehart (1977, 1990) showed that economic development and value change are co-existing effects. That is, the process of economic and technological development triggers changes in individuals' basic values and beliefs (Inglehart & Welzel, 2005). Prior sociological research has shown that early market capitalism resulted in what sociologist Max Weber called the “disenchantment of the world,” stressing rationality and functional utility (Weber, 1978). Following Weber (1978), Inglehart (1977, 1990) argued that industrialization leads to a shift from traditional to secular-rational values. In advertising, rationality and functional utility is reflected in a predominance of cognitive responses that reflect product application, product performance, and benefits that provide functional value. However, later forms of capitalism (or “post-industrialization”) result in a postmodern society and “re-enchantment” and a shift toward post-materialist, emotional values (Firat & Venkatesh, 1995; Inglehart, 1977, 1990; Jenkins, 2000; Ritner, 2005), where hedonic, emotional, and imaginative ads become more important. In other words, as markets mature, consumers take functional features for granted, that is, they know when a product works and are less impressed by the functional attributes displayed in the ads. Thus, they focus on deriving a positive affect from the experiential ad components and become subject to an experiential route of persuasion (Pine & Gilmore, 1999). Indeed, in DMs, where practically all prior ad research has been conducted, a shift from the functional toward the more experiential communications has been reported over the years (Schmitt, 1999; Schmitt, Rogers, & Vrotos, 2003).

However, what about the consumers in EMs? We propose that consumers still primarily respond to functionality because these markets are in earlier stages of capitalism and market development. During the early stages of market development, consumers are more concerned about fulfilling basic rather than high-order needs. Basic needs closely relate to the functional aspects of products, whereas higher-order needs can be fulfilled via the sensory and emotional aspects of products (e.g., aesthetics and self-expression). Finally, consumers in EMs often lack participation in a global consumer market place and are thus less experienced; they are still learning about products and brand differentiation. In sum, we would expect that consumers in EMs are most persuaded by functional advertising communications and engage in cognitive processing, which is subject to a functional route to persuasion. Accordingly, our overall hypothesis can be stated as follows:

In DMs, the experimental route (with experiential messages influencing affect) best describes the advertising process of persuasion. However, in EMs, the functional route (with functional messages influencing cognition) best describes the process of persuasion.

Thus far, we have discussed the relation between functional and experiential aspects of an ad on cognition and affect. However, it is not only ad components per se (functional versus experiential) that influence cognitive and affective ad processing. In addition, ads contain, in their execution styles, certain socio-cultural appeals that are also likely to affect ad processing as well. These socio-cultural ad appeals, being tied to different social and cultural contexts, may result in differential effects between DMs and EMs. Prior social and cultural research has identified several key socio-cultural constructs that have been shown to affect a broad range of consumer behavior. These constructs include a perceived reference dimension (local versus global culture) (Ritzer, 1993), an innovativeness dimension (modern versus traditional culture) (Inglehart, 1997), and, most importantly, a group-related dimension (individualism versus collectivism) (Hofstede, 1980). We next offer some tentative predictions regarding the effects of socio-cultural ad appeals on affect and cognition in general, and how such effects may vary across DMs and EMs.

2.3. Socio-cultural ad appeals and their effect across markets

Based on prior conceptualizations of socio-cultural appeals and on prior research, we expect that ads that appear to connect to a global community rather than a particular culture, ads that appear to be modern in their appeals rather than traditional, and ads that are individualistic rather than collectivistic will result in increased or decreased cognitive and/or affective processing. Most importantly, we expect that these socio-cultural appeals affect cognition and affect DMs and EMs differently.

Regarding the global versus local reference dimension, as part of his work on economic development and cultural change, Norris and Inglehart (2009) recently stressed the role of communications, arguing that in the 21st century, cultural change is driven by information that transcends local communities and national borders and can be characterized as cosmopolitan and global in nature. Global communications represent a global consumer culture that includes symbols and messages that are universally understood by a global community (Ritzer, 1993; Watson, 1997). Advertising contributes to the global consumer culture through what Alden, Steenkamp, and Batra (1999) have called “global consumer culture positioning” (GCCP) in contrast to “local consumer culture positioning” (LCCP) (see also Ford, Mueller, &
In addition to local versus global appeal, another key socio-cultural ad dimension is traditional versus modern appeal (Mueller, 1987). This dimension refers to the perceived innovativeness of a communication (Kunz, Schmitt, & Meyer, 2010). That is, does the ad follow ideas that have existed for a long time, or is it using new ideas and ways of thinking? Traditional ad appeals use themes that look back to the past: they are classic, historical, antique, legendary, time-honored, long standing, venerable, and/or nostalgic (Pollay, 1983). Modern ad appeals, on the other hand, look into the future and include themes that are contemporary, modern, new, improved, progressive, advanced, introducing, and/or announcing (Pollay, 1983).

As modern appeals are associated with “hard-sell” advertising and westernized culture (Chiou, 2002; Lin, 2001; Mueller, 1987), they should impact cognition. However, images of modernity are often multi-sensory, vibrant, and exciting, and thus should also impact affect. Thus, modern appeals should also generate a stronger affective response than traditional appeals.

Finally, ads use individualistic versus collectivistic ad appeal (Zhang, 2010). Dating back to the seminal work by Hofstede (1980), individualism versus collectivism refers to the degree to which individuals are integrated into groups. In individualist societies, the ties between people are loose and are motivated by individual goals. In collectivist societies, people are integrated into strong, cohesive in-groups and motivated by group goals. Ads with individualistic appeal refer to individual aspirations and goal achievement. Ads with collectivist appeals are culturally grounded; thus, they present the social contexts of family, neighborhoods, and friends. Because these ads refer to individual plans and goal-achievement, the more individualistic the ad appeal, the stronger the impact on cognition is. In contrast, the impact on affect should be the opposite, and accordingly, more collectivist ad appeals (displaying groups, friends, children and family) should positively impact affect.

Will there be any differences in the impact of these socio-cultural ad appeals on cognition and affect between DMs and EMs? Given the lack of specific prior research, we must theorize to address this question; therefore, our predictions must be tentative. We propose that there will be differences on all three ad appeal dimensions. Specifically, in DMs, a more global, modern, and individualistic ad appeal should impact affective responses rather than cognitive responses. This is because consumers are used to such messages and to global and modern products for individual use. Therefore, as consumers are unlikely to derive new functional benefits from the products, the consumers are looking for experiences and may enjoy the global, modern, and individualist ad appeal and execution, which is relevant to their life in developed societies and which, as a result, make the brand attractive. In contrast, in EMs, we expect global, modern, and individualistic ad appeals to impact cognition. A global and modern life and lifestyle with individualistic opportunities is what consumers in EMs are striving for, seeking a “passport to global citizenship” (Strizhakova, Coulter, & Price, 2008). Therefore, they will find such messages cognitively appealing in that they provide understanding, credibility, and relevance for the transnational modern-society as well as an individualistic lifestyle, to which they aspire and which are portrayed in these ads.

In DMs, socio-cultural appeals that are global, modern, and individualistic are more likely to influence affect, while in EMs, these socio-cultural appeals are more likely to influence cognition.

3. Data

Our study uses a set of 256 television commercials that were tested by our sponsoring multinational FMCGs corporation in 23 countries, including 17 emerging and 6 developed markets. In total, there are 165 commercials tested in emerging countries and 91 in developed countries. See Table 1.

### 3.1. Country description

Our classification of countries into EMs and DMs is based on two dimensions: the Human Development Index (HDI) (UNDP, 2010) and Inglehart’s (1997) materialist–postmaterialist values priorities. The HDI is a composite score that measures a country’s well-being. Worldwide, the scores, computed based on life expectancy, knowledge and education, and standard of living measures, vary between zero (low HD) and one (high HD) (UNDP, 2010). The HDI scores for the 23 countries (obtained from www.hdr.undp.org/en/media/HDR_2010_EN_Tables_rev.xls.) are reported in Table 2 (last column).

The materialist–postmaterialist values are measured by Inglehart’s (1997, p. 108) 12-item index. For our analysis, we use data collected in the most recent wave 5 (2005 to 2007) of the World Values Survey (WVS; available from http://www.worldvaluessurvey.org/). For each country, the WVS data report the percentage of respondents that fall within each of six materialist–postmaterialist categories (where zero indicates purely materialist and five purely postmaterialist). Table 2 (columns 2 to 7) reports such data for 20 countries in our study. There were no data collected for Pakistan, Philippines, and Saudi Arabia.

### Table 1

<table>
<thead>
<tr>
<th>Emerging countries</th>
<th>Number of ad tests</th>
<th>Developed countries</th>
<th>Number of ad tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>9</td>
<td>Australia</td>
<td>3</td>
</tr>
<tr>
<td>Brazil</td>
<td>3</td>
<td>France</td>
<td>27</td>
</tr>
<tr>
<td>Chile</td>
<td>4</td>
<td>Germany</td>
<td>2</td>
</tr>
<tr>
<td>China</td>
<td>14</td>
<td>Italy</td>
<td>21</td>
</tr>
<tr>
<td>India</td>
<td>46</td>
<td>Netherlands</td>
<td>12</td>
</tr>
<tr>
<td>Indonesia</td>
<td>5</td>
<td>UK</td>
<td>26</td>
</tr>
<tr>
<td>Mexico</td>
<td>3</td>
<td>Total # of ad tests</td>
<td>91</td>
</tr>
<tr>
<td>Morocco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total # of ad tests</td>
<td>165</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The survey consists of representative samples from each country population aged 18 and older, with sample sizes between 902 in Argentina and 2,785 in South Africa. For measuring value priorities, the survey presents respondents with a list of 12 societal goals (e.g., survival and self-expressive goals) and asks them to choose their most and second-most important ones. This procedure delivers, for each respondent, six separate classifications as either purely materialist (scored 0), mixed (1–6), or purely post-materialist (scored 5). See Inglehart (1997, Chapter 4) for more details.
To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia.

Table 2 (column 8) reports the factor scores for the 23 countries.

Fig. 1 maps the 23 countries on the HDI and materialist–postmaterialist dimensions. Using hierarchical cluster analysis (Ward’s method in SPSS), we obtain three groups of countries based on their proximity in the figure and based on a scree plot of the percentage of variance explained by the clusters. The first cluster (located in the upper right of the figure) is composed of postmaterialist, developed societies, including Australia, France, Germany, Italy, Netherlands, and the UK (average HDI = 0.880, average postmaterialist score = 1.24). The second cluster (located in the lower left of the figure) is materialist) is composed of the less developed EMs, including China, India, Indonesia, Morocco, Pakistan, Philippines, Russia, South Africa, Thailand, and Vietnam (average HDI = 0.602, average postmaterialist score = −0.94). The third cluster (located in the middle of the figure) is made up of mixed-type EMs and may be interpreted as transitional economies. Thus, the cluster includes Argentina, Brazil, Chile, Mexico, Poland, Saudi Arabia, and Turkey (average HDI = 0.747, average postmaterialist score = 0.06). Because of the low number of observations in the third cluster (N = 45 ads), we pool the countries in clusters two and three into a single cluster of EMs. In the following, we will refer to the countries in the first cluster as DMs and those in the combined cluster as EMs.

To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia. To index the countries on materialist–postmaterialist values, we factor-analyze the WVS data in Table 2 and obtain a single factor that explains 70.5% of variance (the second factor has an eigenvalue of 1.12 and was dropped to keep the solution parsimonious). To impute the missing values, we regress the country factor scores against the country’s 2006 GDP per Capita and Life Expectancy at Birth and use the resulting equation to predict the scores for Pakistan, Philippines, and Saudi Arabia.
as across markets, because each of the brands advertised has global positioning.

Our unit of analysis is the commercial. Each commercial is measured on two sets of variables. The first set contains aggregate consumer response data measuring consumers’ cognitive, affective, and conative responses to the commercial and is provided by the research institute. The second set contains experts’ judgments of the commercials on various functional, experiential, and socio-cultural dimensions. We now discuss the details of each set of variables.

3.2.1. Consumer response data

This dataset includes the aggregate results of 256 ad tests. Each test is conducted using a sample of 150 consumers who are representative of the country where the test is conducted in terms of gender, age, and socio-economic profile. Thus, the combined dataset represents a worldwide sample of more than thirty-eight thousand consumers. All data are indexed against country norms, where a score of 100 on any particular ad response measure indicates average performance in the country. A score greater (lower) than 100 indicates above (below) average performance in the country. The advantage of such data normalization is that the data from different countries are comparable and there is no “country fixed-effect.”

Consumer responses to advertising were assessed through various measures related to cognitive, affective, and conative responses to advertising. Although not derived from specific academic literature, these measures represent the result of years of practice in the field and have been used repeatedly worldwide. Cognitive response (labeled as “COG”) is measured by five items: 1) ease of understanding the ad (which we label as “Understanding”); 2) credibility of the ad (“Credibility”); 3) relevance of the ad (“Relevance”); 4) degree of differentiation of the ad from others (“Differentiation”); and 5) linkage between the ad and the brand advertised (“Brand identification”). The five measures have high internal consistency, with Cronbach’s alpha equal to 0.89. Affective response (labeled as “AFF”) is measured by two items: 1) enjoyment of the ad (“Enjoyment”) and 2) the attractiveness of the brand in the ad (“Brand attractiveness”). These two measures are internally consistent (Cronbach’s alpha = 0.86).

We use exploratory and confirmatory factor analyses to assess the discriminant validities of the cognitive and affective constructs. The results show that a two-factor solution (with varimax rotation) explains 74% of the variance in the data (41% is captured by the cognitive factor and the remaining 33% is captured by the affective factor). Similarly, a two-factor confirmatory factor analysis (CFA) on the cognition and affect indicators resulted in a significantly superior fit than a single-factor CFA model of all response measures ($\Delta \chi^2 = 110.95, p < 0.001$). All loadings from the two-factor CFA model are significant and large ($p < 0.001$) with Bentler’s comparative fit index (CFI) equal to 0.905 and standardized root mean square residual (SRMR) equal to 0.05. Both fit values are reasonable based on Hu and Bentler’s (1999) cutoff criteria: SRMR is lower than the cutoff value of 0.08, and CFI is close to the cutoff value of 0.95. Table 3 reports the standardized results of the CFA analysis.

Finally, conative or behavioral response is measured by the ability of the ad to persuade consumers to buy the product advertised (purchase intention). We label this variable “PI.” The Appendix lists the set of questions asked by the research institute to measure consumer responses to the commercial.

3.2.2. Experts’ judgment data

Two knowledgeable experts (one senior manager from the sponsoring multinational firm and one co-author) evaluate the 256 TV commercials on more than one hundred measures using a coding scheme we developed.6 In our study, we only use the items that pertain to the evaluation of the commercials on functional (Abernethy & Franke, 1996), experiential (Brakus et al., 2009; Holbrook & Hirschman, 1982; Schmitt, 1999), and cultural (Chiou, 2002; Mueller, 1987; Okazaki et al., 2010) dimensions. See the Appendix for details.

The functional aspects are measured by five indicators that capture the degree to which the commercial focuses on (1) product attributes (labeled as “ATT”); (2) product applications (“APP”); (3) product performance (“PERF”); (4) product benefits (“BEN”); and (5) price/value (“VAL”). Expert judges also evaluate how functional the commercial is overall (“FUNC”) on a four-point scale (1 = not at all functional to 4 = strongly functional). The experiential aspects are measured by four formative indicators that capture the degree to which the commercial appeals to (1) sensory elements (“SEN”); (2) feelings and emotions (“FEEL”); (3) imagination and mental stimulations (“IMAG”); and (4) behaviors and actions (“BEH”). Expert judges also evaluate how experiential the commercial is overall (“EXP”) on a four-point scale (1 = not at all experiential to 4 = strongly experiential). We use three measures for the socio-cultural aspects of a commercial. The measures capture the extent to which the ad has (1) a traditional or modern appeal (“TM”); (2) a local or global appeal (“LG”); and (3) an individual or community appeal (“IC”).

The two expert judges are given all the television commercials with the scripts in the original language and a back-translation in English. After evaluating the commercials independently, the two judges met and compared their codings. We use the procedure suggested by Rust and Cooli (1994) to assess the inter-judge reliability of the data. Specifically, we compute the average reliability value separately for the three-category variables (local/global, traditional/modern, and individualistic/collectivistic) and four-category variables (product attributes, product application, product performance, functional benefits, functional value, sensory elements, feelings and emotions, imagination and mental stimulation, and behaviors and actions) across countries. For the three-category variables, the portion of interjudge agreement is equal to 0.84, which corresponds to a proportional reduction in loss (PRL) of 0.87 (Rust & Cooli, 1994, p. 8). For the four-category variables, the portion of agreement is equal to 0.80, which corresponds to a PRL of 0.86 (Rust & Cooli, 1994, p. 10). As the PRL is comparable to Cronbach’s alpha (Rust & Cooli, 1994), both PRL values indicate a satisfactory inter-judge reliability (Nunnally, 1978). Finally, the judges manage to resolve all conflicts and the agreed-upon coding is merged with the consumer response data, which we use for the empirical analysis.

Table 4 reports the means and standard deviations of all the measures and their correlations.

---

5 It is important to note that the actual coding was performed independent of our study to suit the research goals of the multinational firm. The idea for the present research and the permission to use the data came much after the coding stage. Thus, during the coding stage, neither of the two coders (i.e., the senior manager and the co-author) was aware of the research goals of this paper.

Table 3

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Factor loadings</th>
<th>Error variances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affect</td>
<td>Cognition</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>0.95</td>
<td>0.10</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td>Relevance</td>
<td>0.78</td>
<td>0.39</td>
</tr>
<tr>
<td>Brand identification</td>
<td>0.65</td>
<td>0.57</td>
</tr>
<tr>
<td>Differentiation</td>
<td>0.89</td>
<td>0.21</td>
</tr>
<tr>
<td>Understanding</td>
<td>0.62</td>
<td>0.61</td>
</tr>
<tr>
<td>Credibility</td>
<td>0.90</td>
<td>0.19</td>
</tr>
<tr>
<td>Factor variance</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Factor correlation</td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

* All the factor loadings and error variances are significant ($p < 0.05$).
4. Model

Our conceptual model relating consumer responses to the experiential and functional aspects of the ads, as well as to the socio-cultural ad appeals, is shown in Fig. 2. It is consistent with the general advertising model described earlier. Following the advertising persuasion process, we assume a forward recursive flow of effects from ad aspects through cognitive and affective responses to intended behavior. Working backward, we assume that purchase intent (persuasion) depends directly on two factors: cognition and affect. These two factors, in turn, depend on the functional and experiential aspects of the ad, as well as on the socio-cultural ad appeals. Note that the functional and experiential aspects are endogenously determined by their respective formative indicators, whereas the socio-cultural appeals are treated as exogenous variables.

For model estimation, we measure cognition by the mean of its five indicator variables: understanding, credibility, relevance, differentiation, and brand identification. We also measure affect by the mean of its two indicators, enjoyment and brand attractiveness. Due to the limited sample size, the use of the mean instead of the individual indicators is necessary for the reliable estimation of the model parameters.7

Let i denote commercial l = 1, 2, ..., 256, and let g = 1 ( = 2) denote whether the commercial is tested in an emerging (developed) country. The model, shown in Fig. 2, then simplifies to the following multigroup, simultaneous equation model:

\[
\begin{align*}
\text{FUNC}_i &= \gamma_{01} + \gamma_{11} \text{ATT}_i + \gamma_{21} \text{APP}_i + \gamma_{31} \text{PER}_i + \gamma_{41} \text{BEN}_i + \gamma_{51} \text{VAL}_i + \epsilon_{i1}, \\
\text{EXP}_i &= \gamma_{02} + \gamma_{12} \text{SEN}_i + \gamma_{22} \text{FEEL}_i + \gamma_{32} \text{IMAG}_i + \gamma_{42} \text{BEH}_i + \epsilon_{i2}, \\
\text{COG}_i &= \gamma_{03} + \gamma_{13} \text{FUNC}_i + \gamma_{23} \text{EXP}_i + \gamma_{33} \text{LC}_i + \gamma_{43} \text{TM}_i + \gamma_{53} \text{IC}_i + \epsilon_{i3}, \\
\text{AFF}_i &= \gamma_{04} + \gamma_{14} \text{COG}_i + \gamma_{24} \text{EXP}_i + \gamma_{34} \text{LC}_i + \gamma_{44} \text{TM}_i + \gamma_{54} \text{IC}_i + \epsilon_{i4}, \\
\text{PI}_i &= \gamma_{05} + \gamma_{15} \text{AFF}_i + \gamma_{25} \text{EXP}_i + \gamma_{35} \text{IC}_i + \epsilon_{i5}, \\
\epsilon_i &= \epsilon_{i1} \epsilon_{i2} \epsilon_{i3} \epsilon_{i4} \epsilon_{i5},
\end{align*}
\]

where the \(\gamma\) parameters are regression coefficients to be estimated and \(\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}, \epsilon_{i4}, \epsilon_{i5})\) is a vector of error terms that follows a multivariate normal distribution with a zero mean vector and covariance matrix \(\Psi\). There are two covariance elements of interest. The first, which we denote by \(\Psi_{12}\), is the covariance between \(\text{FUNC}\) and \(\text{EXP}\). This covariance captures the correlation between the extent to which an ad is functional or experiential. The second is the covariance between \(\text{COG}\) and \(\text{AFF}\) and is denoted by \(\Psi_{34}\). This covariance captures the correlation between the cognitive and affective responses. In Fig. 2, \(\psi_{12}\) is represented by the arc connecting \(\text{FUNC}\) and \(\text{EXP}\), and \(\psi_{34}\) is represented by the arc connecting \(\text{COG}\) and \(\text{AFF}\).

There are a few observations regarding the system of equations in Eq. (1). First, because the evaluation of the extent to which an ad is functional or experiential is made by experts, the relationship between \(\text{FUNC}\) and \(\text{EXP}\) and their respective formative indicators is obviously invariant across emerging and developed countries. Second, we do not specify country-specific fixed effects because our data are indexed against country norms (i.e., the data are “mean-centered” by country). Third, the system of equations in Eq. (1) reduces to an aggregate model if the parameters are invariant across groups. We test for such a specification in our empirical analysis.

5. Empirical results

We use our data to estimate the simultaneous system of equations in Eq. (1) with Proc Tcalis in SAS. We specifically estimate two models: an aggregate model that constrains the parameters to be invariant across EMs and DMs, and a multigroup model that allows the parameters to vary across EMs and DMs. We use the latter model to examine if and how the relationship between ad responses and functional and experiential aspects, as well as the socio-cultural appeals, varies across EMS and DMs.

We obtain log-likelihoods of \(-1483.45\) and \(-1455.49\) for the aggregate and multigroup models, respectively. Thus, the multigroup simultaneous equation model has a significantly better fit than the aggregate model (\(\Delta AIC=55.92, p<0.001\)). We arrive at the same conclusion using Akaike’s (1974) information criterion (AIC), which penalizes for over-parametrization: the multigroup model has a lower AIC than does the aggregate model (\(AIC=3014.97\) versus \(AIC=3032.89\), respectively). These results suggest that the drivers of ad performance significantly vary across the two groups of countries.

We now discuss the details of our empirical results by first describing the aggregate results and then the group-level results.

5.1. Aggregate results

We first report the results relating functional and experiential advertising to their respective antecedents and then the results relating these two variables and the socio-cultural appeals to consumer responses. We do so because the latter results are hypothesized to vary across groups of countries, whereas the former results are invariant.

Constraining the model parameters to be invariant across emerging and developed countries, we obtain the following estimates for the first two equations in the simultaneous system of equations in Eq. (1), where parameters in boldface are significant at \(p<0.05\).

\[
\begin{align*}
\text{FUNC}_i &= -31 + 31 \text{ATT}_i + 30 \text{APP}_i - 29 \text{PER}_i + 21 \text{BEN}_i + 0.08 \text{VAL}_i, \\
\text{EXP}_i &= -29 + 33 \text{SEN}_i + 38 \text{FEEL}_i + 28 \text{IMAG}_i + 18 \text{BEH}_i.
\end{align*}
\]

The error standard deviation estimates are 0.41 for the \(\text{FUNC}\) equation and 0.5 for the \(\text{EXP}\) equation. The corresponding R-squared values are, respectively, 0.76 and 0.53, which indicate very good fit. The correlation between the two errors is \(-0.01\) and insignificant. This means that the experts’ evaluations of the extent to which the ads are functional or experiential are independent after controlling for the ad values on the explanatory variables in Eq. (2).

The results in Eq. (2) indicate that when judging the extent to which an ad is functional, experts are more influenced by the degree to which the commercial focuses on product attributes, applications, performance, and benefits than on price/value. Similarly, ad appeals to sensory elements, feelings, and imaginations have more influence on expert judgment of the extent to which the ad is experiential than does appeals to behaviors.

The estimates for the consumer cognitive responses, \(\text{COG}\), \(\text{AFF}\), and \(\text{PI}\), are reported in the top panel in Table 5. Parameters in boldface are significant at the \(p<0.05\) level, and the underlined parameters are significant at \(p<0.1\). All other parameters are insignificant. Note that the parameter estimates can be compared across equations as all three consumer responses (\(\text{COG}\), \(\text{AFF}\), and \(\text{PI}\)) are measured on the same scale. Thus, the results of the aggregate model in Table 5 indicate that functional advertising significantly impacts cognition (\(\psi=1.63, p<0.05\)). Similarly, experiential advertising significantly impacts affect (\(\psi=1.45, p<0.05\)). However, affect is also significantly related to functional advertising (\(\psi=2.17, p<0.05\)). As we discuss below, this effect may be due to aggregation effects (i.e., the pooling of the data across emerging and developed countries). Among the socio-cultural ad appeals, the
local/global variable significantly impacts both cognition and affect (respectively, $\beta = 1.34$ and $\beta = 1.55$; both $p$-values are less than 0.05), whereas the traditional/modern variable significantly impacts cognition only ($\beta = 1.29$, $p < 0.05$). Thus, global ads are likely to lead to higher cognitive and affective responses from consumers, whereas modern ads appear to have a greater impact on cognitive responses. Finally, affect has a relatively larger impact on purchase intention than cognition, even though both variables are significant (respectively, $\beta = 0.70$ and $\beta = 0.33$; both $p$-values are less than 0.05).

To quantify the relative importance of functional advertising and experiential advertising on persuasion, we compute their total effects on persuasion. For example, using the parameter estimates in Table 5 (top panel), the total effect of functional advertising on persuasion is $2.06 (= 1.63 \times 0.33 + 2.17 \times 0.70)$, and the total effect of experiential advertising is $1.19 (= 0.53 \times 0.33 + 1.45 \times 0.70)$. Thus, in the aggregate, functional advertising has a relative importance of 0.63 and is therefore a relatively more important driver of persuasion. These results are reported in Table 6 (first row).

In sum, the aggregate results suggest that (i) functional advertising impacts both cognition and affect, but experiential advertising impacts only affect, and that (ii) functional advertising appears to be a relatively more important driver of persuasion than experiential advertising.

5.2. Multigroup results

Because the extent to which an ad is functional or experiential is judged by experts, the relationships between FUNC and EXP and their respective formative indicators should not vary across EMs and DMs. We already discussed these relationships in the context of the aggregate results. We now focus on examining how the relationship between consumer responses and ad aspects and appeals vary across the two groups of countries, first for DMs and then for EMs.

5.2.1. DMs results

The second panel in Table 5 (upper part) reports the estimates for the simultaneous system of equations in Eq. (1) for DMs. As noted above, the parameter estimates of the FUNC and EXP equations are identical to those reported in Eq. (2) and are, therefore, omitted from the table.

The estimation results for DMs show that cognition is significantly determined by whether the ad is local or global ($\beta = 1.20$, $p < 0.1$), but it is not significantly impacted by whether the ad is functional or experiential. Thus, in DMs, global ads seem to have greater impact on cognitive responses than local ones. The estimation results also show that affect is significantly impacted by experiential advertising ($\beta = 2.71$, $p < 0.01$) and, to a lesser degree, by functional advertising ($\beta = 1.99$, $p < 0.1$). Finally, purchase intent is significantly related to affect ($\beta = 0.98$, $p < 0.05$), but not to cognition. The results in Table 6 (second row), which report the total effects of functional advertising and experiential advertising, suggest that the latter is a relatively more important driver of persuasion than the former. The relative importance of experiential advertising is 0.57.

These findings indicate that, in DMs, both experiential and functional advertising significantly impact persuasion, but the former is a relatively more important driver of persuasion than the latter. Experiential advertising communications produce affective responses which, in turn, impact purchase intention. To be effective, advertising should focus more on stimulating sensations, feelings, imagination, behaviors, and lifestyles.

5.2.2. EMs results

The second panel (lower part) in Table 5 reports the estimates for the simultaneous system of equations in (1) for EMs. The estimation results indicate that functional advertising significantly impacts both cognition and affect (respectively, $\beta = 2.45$ and $\beta = 2.34$; both $p$-values are less than 0.05), whereas experiential advertising impacts neither of these responses. The results also indicate that the local/global appeal has a significant impact on cognition ($\beta = 1.42$, $p < 0.1$). Purchase intent is also significantly related to both cognition and affect (respectively, $\beta = 0.35$ and $\beta = 0.62$; both $p$-values are less than 0.05).

The results in Table 6 (third row) suggest that, in EMs, functional advertising plays a relatively more important role in persuasion than does experiential advertising (relative importance $= 0.72$).
Functional advertising seems to jointly impact both cognition and affect. Thus, to be effective, advertising communications in EMs should focus more on functional and global elements than on experiential aspects.

5.2.3. EMs versus DMs comparison

Thus far, our analysis has focused on assessing the impact of functional and experiential advertising and socio-cultural variables on persuasion in EMs and DMs without assessing whether their differential effect is statistically significant. Following Steenkamp, van Heerde, and Geyskens (2010), we now test whether these effects vary significantly across DMs and EMs. The results of these tests are indicated by superscript “a” in the second panel of Table 5.

These results show that the effect of functional advertising on cognition is significantly different across EMs and DMs (p<0.05). The results also show that the effect of experiential advertising on affect is significantly different (p<0.05). Finally, the impact of modern (versus traditional) socio-cultural appeal on affect differs significantly across EMs and DMs (p<0.05). All the remaining parameters are not significantly different across EMs and DMs (p>0.10). These findings are consistent with our two overall hypotheses: experiential messages influence affect in DMs, and functional messages influence cognition in EMs. Furthermore, modern (versus traditional) socio-cultural appeals influence affect in DMs.

In sum, the aggregate analysis suggests that experiential advertising has impact on affect, whereas functional advertising can impact both cognition and affect. However, these results suffer from aggregation bias that ensues from pooling the data across EMs and DMs. Specifically, for DMs, the multigroup analysis suggests that (1) functional and experiential advertisings impact persuasion only through their effect on consumer affective responses, and (2) cognition has no impact on persuasion. In contrast, for EMs, functional advertising appears to impact both cognition and affect, the two significant
drivers of purchase intent. Experiential advertising, however, has no impact on consumer responses. Finally, the constrained multigroup analysis shows significant differential effect of functional advertising, experiential advertising, and traditional/modern appeals on consumer responses across EMs and DMs.

6. Discussion

Using an extensive data set from an FMCGs company of 256 television commercials for cleaning brands from 23 countries around the world, we find important ad processing differences between EMs and DMs.

In DMs, experiential advertising significantly impacts affect and does not impact cognition. Functional advertising also impacts affect, albeit to a lesser degree. In contrast, in EMs, functional advertising significantly impacts cognition and affect. Both cognition and affect are significant drivers of purchase intent. However, in EMs, experiential advertising has no significant impact on cognition or affect. Thus, in DMs, the experiential route is a more important driver of persuasion; however, the functional route is the key driver of persuasion in EMs. Importantly, these effects are significantly different across DM and EM countries. This supports our overall hypothesis: in DMs, the experiential route best describes the advertising process of persuasion, whereas in EMs, it is the functional route that best describes ad persuasion. Our results also show that, unexpectedly, the functional route influences affect in EMs and, to a lesser degree, in DMs. This result means that functional aspects of the ads, such as product attributes and applications, lead consumers to enjoy the ad and to perceive the brand as attractive. This effect may occur because many products, by their very nature (especially the cleaning products featured herein), offer functionality that creates value, and from this value creation, consumers derive positive affect (Chandy, Tellis, Maclnnis, & Thaivanich, 2001).

Our second overall hypothesis, which states that global, modern, and individualistic ad appeals are more likely to stimulate affect in DMs, whereas in EMs, such appeals are more likely to stimulate cognition, received only partial support. The effects of local/global and modern/traditional ad appeals are largely supported, whereas we find no effects for individualistic versus collectivistic appeals on the persuasion process. As predicted, the global appeal impacts affect in DMs and impacts cognition in EMs: global appeal also has cognitive effects. This impact, however, is not significantly different across the two country groups. With regards to traditional/modern appeals, as predicted, the modern appeal impacts affect in DMs and impacts cognition in EMs. Importantly, this impact is significantly different across the two country groups.

A potential explanation for the lack of effects of the widely studied dimension of individualistic versus collectivistic appeals may be that a truly cultural concept, such as individualism versus collectivism, may become increasingly less relevant in an increasingly globalized world driven by consumer culture. That is, unlike local/global appeals and modern/traditional appeals, which refer to appeals through ad execution styles and, thus, to “consumer culture,” individualistic versus collectivistic appeal refers to genuinely cultural content (individualism versus group). In a globalized consumer world, consumers may, in general, act increasingly more individualistic, and thus, the cultural difference may disappear as consumers are more affected by emerging consumer culture than by century-long cultural traditions.

7. Limitations and future research

Our research reveals important differences in the advertising persuasion process between EMs and DMs; however, the results are also subject to several limitations. First, the paper uses a dataset that includes only one product category (household cleaners). Future research should include other categories and test whether the results generalize to other categories, for example, to higher involvement products, such as fashion or automotive brands. A second limitation concerns the medium investigated here—television. Future research should concentrate on non-television communications and investigate whether the same persuasion-processing differences between markets can be found for other media as well. Finally, although the sample used included several ads from a large set of countries, the number of observations and countries was insufficient to investigate further differences among emerging countries. In particular, the limited number of observations related to transitional economies did not allow us to compare transitional economies, such as Argentina, with less developed economies, such as China or India. Future research should deepen our understanding of the advertising persuasion process in EMs by including additional ad dimensions and by categorizing EMs along other pertinent constructs, such as ethnicity, history, and religion.

Acknowledgments

The authors thank a large fast-moving consumer goods multinational company that provided the dataset. In particular, they thank the two managers who actively supported the project and the
### Appendix. Consumer response and experts’ measures

<table>
<thead>
<tr>
<th>Dimension measured</th>
<th>Items</th>
<th>Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consumer response data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>1. Understanding: How easy was it to understand what was going on in the advertisement?</td>
<td>1. Four-point scale from “Very hard” to “Very easy”</td>
</tr>
<tr>
<td></td>
<td>2. Credibility: How strongly do you agree or disagree that what the advertisement puts across about brand X is believable?</td>
<td>2. and 3. Five-point scale from “Strongly disagree” to “Strongly agree”</td>
</tr>
<tr>
<td></td>
<td>3. Relevance: If you were buying a household cleaner, how relevant would the points made in the advertisement be to you?</td>
<td>4. Four-point scale from “Not at all relevant” to “Very relevant”</td>
</tr>
<tr>
<td></td>
<td>4. Differentiation: How different is this advertisement from others that you have seen?</td>
<td>5. Five-point scale from “It could have been for almost anything” to “You couldn’t fail to remember it was for brand X”</td>
</tr>
<tr>
<td>Affect</td>
<td>1. Enjoyment: How much would you enjoy watching this advertising each time you see it on television?</td>
<td>1. Five-point scale from “Not at all” to “A lot”</td>
</tr>
<tr>
<td></td>
<td>2. Attractiveness: How much is the ad able to increase the appeal of brand X?</td>
<td>2. Five-point scale from “Much less appealing” to “Much more appealing”</td>
</tr>
<tr>
<td>Purchase Intention</td>
<td>1. How will the advertising affect your use of brand X?</td>
<td>Four-point scale from “Makes me less likely to continue using brand” to “Strongly encourage me to continue using brand X”</td>
</tr>
<tr>
<td><strong>Experts’ judgment data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional aspects of ads</td>
<td>To what degree does the ad focus on:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Product attributes (i.e., the formulation or ingredients of the product and its features)?</td>
<td>1. to 5: 1 = Not at all present, 2 = Poorly present,</td>
</tr>
<tr>
<td></td>
<td>2. Product application (i.e., how the product has to be applied or rinsed; example: instructions for use, dosage, implement required)?</td>
<td>3 = Somewhat present, 4 = Strongly present</td>
</tr>
<tr>
<td></td>
<td>3. Product performance (i.e., what the product can do and its cleaning efficacy)?</td>
<td>6. 1 = Not at all functional, 2 = Poorly functional,</td>
</tr>
<tr>
<td></td>
<td>4. Functional benefits (i.e., the advantages for the consumer)?</td>
<td>3 = Somewhat functional, 4 = Strongly functional</td>
</tr>
<tr>
<td></td>
<td>5. Functional value (i.e., value for money or convenience of the product)?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Overall functional (i.e., an ad that includes the above and related characteristics)</td>
<td></td>
</tr>
<tr>
<td>Experiential aspects of ads</td>
<td>To what degree does the ad use or appeal to:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Sensory elements (i.e., colors and exciting visuals, music, touch, smell)?</td>
<td>1. to 4. 1 = Not at all present, 2 = Poorly present,</td>
</tr>
<tr>
<td></td>
<td>2. Feelings and emotions (i.e., all kinds of feelings and emotions, either positive such as joy or negative such as fear)?</td>
<td>3 = Somewhat present, 4 = Strongly present</td>
</tr>
<tr>
<td></td>
<td>3. Imagination and mental stimulation (i.e., thinking in a different, original and innovative way, approaching things from a new angle)?</td>
<td>5. 1 = Not at all experiential, 2 = Poorly experiential,</td>
</tr>
<tr>
<td></td>
<td>4. Behaviors and actions (i.e., physical activities, specific actions, bodily experiences)?</td>
<td>3 = Somewhat experiential, 4 = Strongly experiential</td>
</tr>
<tr>
<td></td>
<td>5. Overall experiential (i.e., an ad that includes the above and related characteristics)</td>
<td></td>
</tr>
<tr>
<td>Socio-cultural ad appeal</td>
<td>The ad:</td>
<td>For all questions: 1 = Has a more local (or traditional or individual…) than global (or modern or group/community…) appeal; 2 = Has an equally local and modern appeal; 3 = Has a more local than modern appeal</td>
</tr>
<tr>
<td></td>
<td>1. Local/global. Has a local or global appeal (local = country specific, connecting with a particular culture, place or area; global = universal or inter-cultural, can travel across different countries without specific need of translation)?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Traditional/modern. Has a traditional or modern appeal (traditional = conventional, following ideas and methods that have been existing for a long time; modern = up-to-date, using or willing to use very recent ideas, fashions or ways of thinking)?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Individualistic/collectivistic. Talks about the individual or a group/community (individual = self, single person and his/her world; group/community = a group of persons such as family, neighborhood, friends)?</td>
<td></td>
</tr>
</tbody>
</table>

director of the marketing research division who made this project possible. The authors also acknowledge Jia Liu for her assistance in the data analysis as well as the two editors of the special issue and the two anonymous reviewers for their comments.

### References


Winning hearts, minds and sales: How marketing communication enters the purchase process in emerging and mature markets

Koen Pauwels a,⁎, Selin Erguncu b, Gokhan Yildirim c

a Marketing at Ozyegin University, Istanbul, Turkey
b Koc University, Istanbul, Turkey
c Management School, Lancaster University, UK

1. Introduction

Both the opportunities and the threats of increasing globalization have created an urgency for companies to succeed in international markets (Burgess & Steenkamp, 2006; Chao, Samiee, Sai, & Yip, 2003). Companies from mature markets strive to win hearts and sales in emerging markets, which will account for most of the economic growth in the coming decades. For example, General Motors and Peugeot struggled to obtain a share of the Chinese market (Bizious & Crawford, 1997; Engario, Kripalani, & Webb, 2001), at least in part due to cultural misunderstandings (Chen, 2001). At the same time, brands from emerging markets, such as Lenovo and Haier, struggle to succeed in mature markets (Pukthuanthong & Roll, 2009) at least in part because they lack a strong emotional connection with their customers (Lindstrom, 2011; Wang, 2008). The opening quotes illustrate the clash between views that marketing principles are universally applicable and observations of different consumers’ responsiveness to marketing communication. Is it truly the case that, in emerging markets, “building consumer hearts and minds” (Kotler & Pfoertsch, 2010) translates into higher sales? Can systematic differences in emerging versus mature markets help us predict how marketing communication enters the purchase process and converts into sales? These are the questions that guide us in this paper.

Despite considerable research on emerging markets, important knowledge gaps remain on whether and how marketers can influence consumer perceptions, attitudes and intentions— all of which we refer to as the “consumer mindset.” While some researchers find that cognitive decision processes are universal across consumers (e.g., Aaker & Maheswaran, 1997; Douglas & Craig, 1997), they leave
open the possibility of substantial differences in the extent to which the process components influence purchase and the power of marketing to affect this process. Such issues are largely unanswered in cross-cultural marketing research, which has focused on country-of-origin effects, consumer perception of local versus global brands (e.g., Batra, Ramaswamy, Alden, Steenkamp, & Ramachander, 2000; Ozsomer, 2012; Steenkamp, Batra, & Alden, 2003) and the content of advertising appeals (e.g., Aaker & Williams, 1998; Han & Shavitt, 1994). Though important, these factors do not address a more general question. Should brand managers focus on moving the needle on different aspects of the consumer mindset in emerging versus mature markets? Recent conceptual papers hint that this may be the case. For example, Burgess and Steenkamp (2011) and Cayla and Arnould (2008) highlight cultural differences in the importance of individual versus group decision making as a key reason for different branding strategies in emerging versus mature markets. What is currently missing is a conceptual model and empirical approach to analyze these differences and provide guidance to marketers aiming to increase brand sales in emerging and mature markets.

We propose that marketing effectiveness differs in the extent to which consumers (1) become aware of marketing communication, (2) are open to change their minds and hearts and (3) change their buying patterns accordingly. These properties may differ from consumer (group) to consumer (group) within a country, but also should systematically differ among consumers from a mature versus an emerging market. If this is the case, conceptual arguments and findings regarding consumer attitudes and behavior based on mature markets may not hold in emerging markets. Key examples include the mandate that brands should be romantic and mysterious ‘love marks’ (Roberts, 2005), and the finding that brand liking is very responsive to advertising and converts strongly into sales (Hanssens, Pauwels, Srinivasan, & Vanhuele, 2010). Based on the three “pillars of institutions” in institutional context theory (Burgess & Steenkamp, 2006; Scott, 2001), we propose that differences in regulative, cultural and economic systems reduce the generalizability of such findings. We analyze the extent of consumer protection as the regulative factor. As a key cultural difference, we focus on Hofstede’s (1980) individualism/collectivism dimension and incorporate income level as the economic factor. Differences among these three systems translate into specific propositions on the marketing responsiveness and sales conversion of consumer mindset metrics.

Our contributions are twofold. First, we provide a unifying conceptual framework to translate consumer differences into observable criteria of market-level mindset metrics. Second, we empirically demonstrate the proposed differences in a longitudinal hierarchical linear model estimated on a unique dataset containing marketing, sales and consumer mindset metrics in Brazil and the U.K. As an initial test of our framework, this empirical study provides novel insights on how marketing enters the purchase process in a major emerging and a major mature country market.

The remainder of this paper moves from the research background and hypotheses to our conceptual framework and effectiveness criteria for mindset metrics. Combining both building blocks, we propose our conceptual framework of how institutional context differences affect mindset metric effectiveness criteria in an emerging versus a mature market. From this framework, we derive specific hypotheses for our empirical setting of a major emerging market (Brazil) versus a major mature market (the U.K.).

### 2.1. Regulative, cultural and economic system differences

As part of the regulative context, consumer protection against poor-quality products appears especially relevant to our study of how consumers respond to marketing communication. Lack of such protection is a key example of an ‘institutional void’ typically found in product markets of emerging countries (Khanna & Palepu, 2010). Beyond the existence of quality and safety regulations, Khanna and Palepu (2010) also ask the following questions: “How do the authorities enforce regulations?”, “What recourse do consumers have against false claims or defective products?”, “Can consumers easily obtain unbiased information about the quality of the goods and services they want to buy?” and “Are there independent consumer organizations and publications that provide such information?” Marketing literature has long demonstrated that quality uncertainty increases consumers’ risk perceptions, which leads them to search for more quality information before purchase (Erdem, Swait, & Valenzuela, 2006; Money, Gilly, & Graham, 1988; Shimp & Bearden, 1982). In contrast, consumers enjoying strong protection may “assume that all brands offered by mainstream retailers deliver the same basic quality” (Hollis, 2010).

As to culture, one of the main issues facing all societies is to define the nature of the relation between the individual and the group (Schwartz, 1999). Researchers have labeled this tension as independent versus interdependent self-construal (Doi, 1986; Markus & Kitayama, 1991), individualism versus collectivism (Hofstede, 1980), separateness versus interdependence (Kagitcibasi, 2005) and autonomy versus relatedness (Schwartz, 1999). Following Hofstede (1980), we use the term “individualism” to identify the relative emphasis on the individual versus the larger social group. People in individualist cultures believe that individual is the most important unit. They are self-oriented, make their decisions based on individual needs and independently pursue their own ideas and preferences. Conversely, people in collectivistic cultures believe that the group is the most important unit. They are group-oriented, their decisions are based on what is best for the group and, identifying with the group and participating in its shared way of life, they find meaning in life largely through social relationships (Hofstede, 1980). Individualism—collectivism is “perhaps the most central dimension of cultural variability identified in cross-cultural research” and has inspired a substantial body of research in marketing (Aaker & Maheswaran, 1997, p. 315). Practical implications for managers are detailed in, e.g., Wang’s (2006) distinction of how L’Oréal should implement different branding strategies in an individualist versus collectivist society.

As to the economic context, the gross domestic product (GDP) per capita or other measures that express purchasing power have long been used as the defining difference between mature versus emerging country markets (World Bank, 2010). Compared to the more complicated human development index of the United Nations, the GDP per capita criterion is easier to use and is more directly relevant to marketing as it focuses on available monetary resources in the country (Burgess & Steenkamp, 2006). When the GDP per capita is low, it is harder for consumers to ‘follow their heart’; no matter how much consumers love a brand, they will not buy it if it is not affordable (Pfeiffer, Massen, & Bombka, 2007). Even for products considered low-ticket in mature markets, price can be an important purchase obstacle for emerging market consumers despite their positive disposition towards the brand (Steenkamp & Burgess, 2002).
2.2. Consumer mindset metrics and their effectiveness criteria

Marketing literature is rich in conceptualization and measurement of consumer mindset metrics, such as communication awareness, brand awareness, brand consideration, brand liking, and brand preference. Although there is consensus that these metrics, in general, help detect and understand the process from brand exposure to purchase (Keller & Lehmann, 2006), debate has raged over which metrics matter (e.g., Lautman & Pauwels, 2009) and over whether the metrics fit into a hierarchical, linear ‘purchase funnel’ (e.g., Palda, 1964) or operate in a parallel fashion, as suggested by neuroscience (e.g., Rose, 1993). Empirical evidence indicates that (1) communication awareness, brand consideration and brand liking metrics substantially improve the predictive power of marketing models (Srinivasan, Pauwels, & Vanhuele, 2010) and (2) parallel impact of such metrics predicts sales better than any hierarchy does (Vakratsas & Ambler, 1999). We maintain these assumptions in our model.

For marketing to effectively change behavior, consumers must become aware of marketing communication, must be open to change their minds and hearts, and, consequently, their buying patterns. The first part refers to the responsiveness of communication awareness1 to marketing communication. The second part refers to the responsiveness of brand attitudes, such as brand consideration and brand liking. Srinivasan et al. (2010) and Hanssens et al. (2010) propose consideration (set inclusion) to represent the ‘cognitive’ dimension, i.e., consumers’ minds, and propose the extent of brand liking to represent the ‘affective’ dimension; i.e., consumers’ hearts. Finally, the third part refers to the sales conversion of communication awareness, brand consideration and brand liking.

Recently, Hanssens et al. (2010) operationalized effectiveness criteria for consumer mindset metrics to capture their (1) responsiveness to marketing, (2) stickiness and (3) sales conversion. First, responsiveness is measured as the elasticity of each mindset metric to marketing, accounting for diminishing returns as the mindset metric runs out of potential to grow (e.g., 99% awareness). Second, stickiness refers to the staying power of a change in the mindset metric in the absence of further marketing effort. It is measured in a regression of the mindset metric on its own past. Finally, sales conversion is measured as the elasticity of brand sales to each mindset metric. Managers are urged to focus on marketing actions that generate a large response in a mindset metric that has high staying power and converts strongly into sales.

2.3. Conceptual framework and hypotheses

Fig. 1 displays our conceptual framework. Starting from differences in regulative protection, individualism and income levels, we propose different responsiveness, stickiness and sales conversion of mindset metrics in an emerging versus a mature market.

Our framework is general, as are the conceptual arguments for our hypotheses. Moreover, we built on previous findings from several emerging and mature markets (e.g., Erdem et al., 2006; Hult, Hurley, Giunipero, & Nichols, 2000; Money et al., 1988; Nicholls, Roslow, & Dubush, 1997), reflecting the view that similarities in institutional context abound among countries within each group (Burgess & Steenkamp, 2006). At the same time, we acknowledge substantial differences within emerging and mature markets. Therefore, any empirical analysis can only provide a partial assessment of the framework and should formulate hypotheses specific to the analyzed markets. We provide a first empirical assessment with the specific institutional context differences between Brazil (a major emerging market) and the U.K. (a mature market). Our interest generates from the distinct differences between the average Brazilian and the average U.K. consumer on the three institutional context dimensions.

First, the Brazilian consumer enjoys less consumer protection against poor quality products than the U.K. consumer. In Brazil, the Consumer Protection Code, which establishes basic consumer rights and sets penalties for infractions, was introduced in 1990 (Pinto, 2002). In the U.K., such regulations were enacted in the 1970s (Beale, 1978), Proteste, the Brazilian Association of Consumer Protection, celebrated its 10th birthday in 2011 (http://www.proteste.org.br/), while Consumer’s Union celebrated its 75th. The U.K. also has a designated government office to address consumer complaints – the Office of Fair Trading, established by the Fair Trading Act of 1973. In contrast, the 1990 Consumer Protection Code in Brazil only establishes the Consumer Protection National System, which loosely combines the country’s and civil society’s efforts, leaving consumer protection “without a specific centralization” (Pinto, 2002, p. 17). Due to regional disparities, lack of resources and commercial pressure, Pinto (2002) concludes that, despite progress, “in several layers of society, citizens still ignore their basic consumer rights” (p. 31). This lack of consumer protection reflects the regulative context in general. LaPorta, Lopez-de-Silanes, Shleifer, and Vishny (1998) score the U.K. 10/10 for efficiency of judicial system against 5.75/10 for Brazil, while the Global Corruption Report (2009) gives U.K. a 7.7/10 and Brazil a 3.5/10 (with 10 meaning ‘highly clean’).

How would these differences in regulative protection affect the consumers’ purchase process? On the one hand, consumers may trust marketing communication less as they enjoy less protection against misleading marketing. However, empirical studies consistently show that consumers in emerging markets pay more attention to marketing communication and trust marketing messages more than their counterparts in mature markets (Eisend & Knoll, 2012; Mindshare, 2011; Möller & Eisend, 2010; Nielsen Media Research, 2009). For instance, 74% of consumers in Latin America (77% in Brazil) trust television advertising compared to only 49% in the EU (Nielsen Media Research, 2009). Similarly, 82% of consumers in Latin America agree with the statement that “by providing information, advertising allows for better consumer choices” compared to 50% in the EU. These recent numbers support the argument that the information function of marketing is higher in emerging markets (Burgess & Steenkamp, 2006). In a caveat emptor (buyer beware) environment, the buyer is the main responsible party for ensuring that product quality meets minimum standards (Andaleeb & Anwar, 1996; Qu, Ennew, & Sinclair, 2005). Concerns to avoid poor quality products induce consumers to attend more to communication regarding the quality of brands (Erdem et al., 2006). Due to the responsiveness to marketing communication, future marketing stimuli will weaken the recall of the current stimulus (Burke & Srull, 1988; Keller, 1987). As a result, increases in communication awareness are more difficult to maintain in the absence of repetition, leading to reduced stickiness. Combining our predictions with the current situation of regulative protection in Brazil and the U.K., we propose the following:

**Hypothesis 1.** For Brazil versus the U.K., communication awareness is (a) more responsive to marketing communication and (b) less sticky.

Second, on Hofstede’s (1980) scale of individualism, Brazil scores 38 and the U.K. 90 out of 100. We posit that individualism affects the responsiveness of brand attitudes to marketing communication. Living in a highly individualist culture, U.K. consumers see themselves as independent and distinct from the group, and accordingly, they place a high value on uniqueness, individual accomplishments and achievement. As a result, they should feel free to change their own brand attitudes substantially based on marketing communication. In contrast, Brazilian consumers see themselves as part of a larger group, and accordingly,
they value connectedness and conformity and are integrated into strong, cohesive in-groups. As a result, they should be less willing to change their attitudes based solely on marketing communication. Instead, their attitude changes mostly derive from social interaction. “If a symbol is to convey meaning, it must be identified by a group [...] and the symbol must convey similar meaning to all within the group” (Grubb & Grathwohl, 1967, p. 24).

The few empirical papers on the subject attest to the notion that marketing communication is less important than social influence for consumers in collectivist cultures. Nicholls et al. (1997) find that Hispanic customers are more influenced by social influences than their Anglo counterparts in the U.S. Similarly, Money et al. (1988) report that consumers in collectivist cultures rely more on interpersonal information exchange or word-of-mouth. Brands that are considered expert and trustworthy are more valuable in collectivist cultures because they help reinforce group identity (Erdem et al., 2006; Johanson, Ronkainen, & Czinkota, 1994). This anchoring of brand attitudes in the group or community also implies that, when a brand does succeed in improving attitudes, this change is rather enduring, i.e., sticky. Therefore, we predict the following hypotheses on the basis of our conceptual framework:

**Hypothesis 2.** With respect to Brazil versus the U.K., the brand attitudes, consideration and liking, are (a) less responsive to marketing communication and (b) more sticky.

Finally, income plays a key role in the conversion of brand liking into purchase. Poorer consumers spend a large part of their income on daily use products (World Bank, 2010). Many economic studies have found that low-income consumers make more rational versus emotional purchase decisions (Jones & Mustiful, 1996). Low-income consumers focus on value and product functionality (Cayla & Arnould, 2008), which drive brand consideration (Kardes, Kalyanaram, Chandrasekaran, & Dornoff, 1993; Roberts & Lattin, 1991). In contrast, brand love is driven by self-brand integration, passion, separation distress, romance, mystery and sensuality (Batra, Ahuvia, & Bagozzi, 2012; Roberts, 2005). High-income consumers have the luxury to buy the brands they love, because they spend a small proportion of their income on consumer products. Moreover, high-income consumers tend to gain greater command of their own information environments and are more likely to rely on their own brand liking in their purchase decisions (Bennett, 1998; Giddens, 1991). Comparing Malaysia with France, Hult et al. (2000) find that consumers in the lower-income country place more importance on tangible attributes, such as price and safety. Given that the per capita gross domestic product (PPP) is $10,800 for Brazil versus $34,800 for the U.K. (World Factbook, 2011), we posit the following:

**Hypothesis 3.** With respect to Brazil versus the U.K., brand liking has a lower sales conversion.

We summarize the conceptual arguments and our specific hypotheses in Table 1.

### 3. Empirical study

Our conceptual framework may be falsified by different data collection methods, including experiments, surveys and purchase behavioral data. Several data providers have measured consumer attitudes at the market or segment level for decades and thus have achieved adequate representation and sample sizes. The resulting metrics (including price image, communication awareness, consideration and liking) predict sales (Hanssens et al., 2010; Lourenço, 2011, Srinivasan et al., 2010, Van Heerde, Gijsbrechts, & Pauwels, 2008). Moreover, managers are encouraged to use such mindset metrics to evaluate the success of their marketing communication actions (Keller & Lehmann, 2006, Pauwels & Joshi, 2011). Despite the benefits of external validity and actionability, these data also have drawbacks as they are not available at the individual consumer level and constructs cannot be manipulated.

The level of analysis is an important choice. Once we move beyond the individual consumer level, we can formulate hypotheses for groups of consumers, whether these are subcultures within a city (e.g., Ackerman & Tellis, 2001), age cohorts within a country (e.g., Inglehart & Baker, 2000) or countries (e.g., Hofstede, 1980; Schwartz, 1999). The latter level of analysis is typical in cross-cultural research and has the benefit of currently available data on marketing communication spending, mindset metrics and sales. Moreover, previous literature has established average levels of individualism/collectivism, consumer protection and income at the country level. These benefits come at a cost, however, as the analysis at the country level masks differences among regions within a country and among consumers within a region.

Our empirical study combines archival sales and marketing information with large-sample survey data on consumer attitudes at the country level for Brazil and the U.K. These markets are of commercial interest because they represent a major emerging and a major mature market, each of which place in the top 10 in the category’s worldwide consumption.

#### 3.1. Data

The dataset contains 72 monthly observations on marketing actions (price, distribution and advertising), sales, and mindset metrics.
(communication awareness, brand consideration and brand liking) for 6 brands in Brazil and 10 brands in the U.K. The operationalizations follow standard practice: sales and prices are expressed in ounces of the product, distribution is all commodity value (ACV) distribution in the country and advertising is measured in gross rating points (GRPs). To control for inflation, we calculate the relative price as brand price divided by category price. The mindset metrics are similar to those in Srinivasan et al. (2010), as detailed in Table 2.

Monthly sample sizes for these mindset metrics exceed 200 in each country, and quota sampling ensures that sampled respondents are representative of the country’s consumers in the category. This characteristic increases the comparability between the emerging market and the mature market sample (Sekaran, 1983) as well as the managerial relevance of our findings.

The data provider requires confidentiality regarding the identity of the personal care category and that of its brands, which are formulated and positioned either for males or for females. Table 3 presents the descriptive statistics on each variable for all advertised brands.

Table 1: Summary of conceptual arguments and findings.

<table>
<thead>
<tr>
<th>Institutional dimension</th>
<th>Theoretical argument</th>
<th>How Brazil differs from the U.K.</th>
<th>Hypotheses for Brazil versus the U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulative</td>
<td>Concerns to avoid poor quality products should lead consumers to attend more to communication on the quality of brands (Ertem et al., 2006)</td>
<td>Brazilian consumers enjoy less consumer protection against poor-quality products than U.K. consumers (LaPorta et al., 1998)</td>
<td>Communication awareness is more responsive to marketing communication [H1a; supported] Communication awareness is less sticky to marketing communication [H1b; supported]</td>
</tr>
<tr>
<td></td>
<td>Increases in communication awareness are harder to maintain in the absence of repetition (Burke &amp; Sruil, 1988; Keller, 1987)</td>
<td>On Hofstede’s (1980) individualism scale Brazil scores 38 and the U.K. 90 out of 100</td>
<td>Brand attitudes are less responsive to marketing communication [H2a; supported] Brand attitudes are more sticky to marketing communication [H2b; not supported]</td>
</tr>
<tr>
<td>Cultural</td>
<td>Collectivism implies marketing communication to be less important than social influence (Money et al., 1988-Nicholls et al., 1997)</td>
<td>The per capita gross domestic product is $10,800 for Brazil versus $34,800 for the U.K. (World Factbook, 2011)</td>
<td>Brand liking has a lower sales conversion [H3; supported]</td>
</tr>
<tr>
<td></td>
<td>When a brand does succeed in improving attitudes, this change is more enduring in collectivist cultures (Johansson et al., 1994)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic</td>
<td>Low-income consumers make more rational versus emotional purchase decisions (Cayla &amp; Arnold, 2008; Jones &amp; Mustalit, 1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High-income consumers are more likely to rely on their own brand liking in their purchase decision (Bennett, 1998; Giddens, 1991)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Variable operationalization.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marketing mix</td>
<td>Average price paid for 1 oz of brand, divided by average price in category</td>
</tr>
<tr>
<td>Relative price</td>
<td>All Commodity Volume (ACV) weighted distribution</td>
</tr>
<tr>
<td>Distribution</td>
<td>Gross rating points (GRPs) of advertising</td>
</tr>
<tr>
<td>Advertising</td>
<td>Volume sales in ounce</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Mindset metrics</td>
<td>“For which of these brands have you seen, heard, or read any advertising in the past 6 months?” (Respondent reads a list of brands, and indicates YES or NO to each)</td>
</tr>
<tr>
<td>Communication awareness (% aware)</td>
<td>% aware is the percentage of respondents indicating ‘YES’ for the particular brand</td>
</tr>
<tr>
<td>Brand consideration (% considering buying)</td>
<td>% consideration is the percentage of respondents indicating ‘YES’ for the particular brand</td>
</tr>
<tr>
<td>Brand Liking (% of liking)</td>
<td>“Please indicate how much you like brand X.” (1: I don’t like at all, 7: I like a lot)</td>
</tr>
</tbody>
</table>

a Measured every month in a stratified national sample, with between 246 and 251 respondents realized each month in Brazil, and between 243 and 249 respondents realized in the United Kingdom.

3.2. Methodology

Our empirical methodology starts from Hanssens et al. (2010): they specify separate regressions for responsiveness of each mindset metric, for stickiness of each mindset metric, and for the sales conversion of the mindset metrics. They also note an important methodological issue, that is, while sales conversion of mindset metrics is likely a characteristic of the consumer decision process in the category (and the country), responsiveness to marketing is likely brand-specific. Thus, we need to account for both country market and brand variation in the coefficients relevant to our hypotheses.

2 We both perform the augmented Dickey Fuller (ADF) test, which has evolution as the null hypothesis, and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test, which has stationarity as the null hypothesis.
Hierarchical linear models (HLMs) are designed to analyze multilevel data (Draper, 1995) and can incorporate heteroskedasticity and dependence. The HLM’s mathematical form enables researchers to investigate the underlying theory about the functional relationship among the variables in each level (Heck & Thomas, 2000). The variance of an outcome variable is partitioned into within and “between” variances, which should increase the precision of estimates. In matrix form, the general specification is:

\[ y = X\beta + Zu + \epsilon \]  

(1)

where \( y \) is an \( n \times 1 \) vector of responses, \( X \) is an \( n \times p \) matrix containing the fixed effect regressors, \( \beta \) is a \( p \times 1 \) vector of fixed effects parameters, \( Z \) is an \( n \times q \) matrix of random effects regressors, \( u \) is a \( q \times 1 \) vector of random effects, and \( \epsilon \) is an \( n \times 1 \) vector of errors.

In our three-level model, time series observations within brands constitute the first level, the brands constitute the second level, and the markets constitute the third level. As a result of this hierarchical structure, the model analyzes the brands in common across countries. We fit HLM by combining fixed and random effects. We allow for random effects at both the market and the brand-within-market levels. We choose the higher likelihood model from between 1) the varying-intercept (random-intercept) model and 2) the varying-intercept and varying-coefficient (random-intercept and random-slope) model. Summing up the random and fixed effects, we derive separate values for the coefficients of interest for Brazil and the U.K.

Responsiveness is the response of each mindset metric to marketing. As do Hanssens et al. (2010), we use the multiplicative model and incorporate diminishing returns by expressing the dependent variable as an odds ratio of the mindset metric (e.g., 60% awareness) and its remaining potential (e.g., 100% − 60% = 40%). The HLM specification is:

\[ y_{ijk} = \alpha + \beta_{yijk} X_{ijk} + \epsilon_{ijk} + \sigma(3)_{k} \]

(2)

where \( y \) is the log of odds ratio \( Y/(100\% − Y) \), \( Y \) is the mindset metric, and \( X \) are the logits of marketing (relative price, distribution and advertising GRPs). The index \( i \) is for time series observations, \( j \) for brands, and \( k \) for markets. \( \sigma(3)_{k} \) is the random intercept for markets \( k \), and \( \sigma(2)_{jk} \) is the random intercept for brand \( j \) and market \( k \). Finally, \( \epsilon_{ijk} \) is the residual error and \( \beta_{yijk} \) are the responsiveness coefficients of interest. As do Hanssens et al. (2010), we run the model separately for each mindset metric (communication awareness, brand consideration, and brand liking).

Stickiness is captured by an autoregressive (AR) process, i.e., regressing each mindset metric on its own lagged value. The stickiness value acts as a multiplier for translating short-term into long-term gain. For stickiness values of 0.9, 0.8 and 0.5, respectively, one multiplies the short-term gain in the mindset metric by 10 \([-1/(1 − 0.9)]\), 5, and 2, respectively, to obtain the long-term gain without any further stimulation. The HLM specification is:

\[ y_{ijk} = \alpha + \beta_{yijk} X_{ijk} + \sigma(2)_{jk} + \sigma(3)_{k} + \epsilon_{ijk} \]

(3)

where \( x \) is the lagged dependent variable and \( \beta_{yijk} \) is the ‘stickiness’ coefficient of interest, which varies across markets and across brands.

We assess sales conversion in a single model in which we allow for each attitude to influence sales (Hanssens et al., 2010; Vakratsas & Ambler, 1999). This also makes it possible to empirically test for, e.g., a higher sales conversion of liking in the U.K. versus Brazil but a lower conversion of communication awareness. The HLM specification is:

\[ y_{ijk} = \alpha + \beta_{yijk} X_{ijk} + \epsilon_{ijk} + \sigma(3)_{k} \]

(4)

where \( y \) is the log of sales volume, \( X \) is the log of each of the 3 mindset metrics, and \( \beta_{yijk} \) are the sales conversion coefficients for each mindset metric.

\[ \text{Table 3} \]

Descriptive statistics for brands present in both markets (ordered by market share) and remaining brands (ordered by market share).

<table>
<thead>
<tr>
<th>M/female sales rank</th>
<th>M1 Mean</th>
<th>M1 SD</th>
<th>M2 Mean</th>
<th>M2 SD</th>
<th>M3 Mean</th>
<th>M3 SD</th>
<th>M4 Mean</th>
<th>M4 SD</th>
<th>M5 Mean</th>
<th>M5 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consideration</td>
<td>66.043</td>
<td>5.768</td>
<td>67.878</td>
<td>6.033</td>
<td>21.304</td>
<td>4.646</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price</td>
<td>0.987</td>
<td>0.022</td>
<td>0.931</td>
<td>0.018</td>
<td>1.154</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td>474.671</td>
<td>174.395</td>
<td>580.884</td>
<td>86.639</td>
<td>189.329</td>
<td>70.529</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ad GRPs per 100 m inhabitants</td>
<td>171.293</td>
<td>280.301</td>
<td>229.043</td>
<td>282.048</td>
<td>7.480</td>
<td>32.502</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consideration</td>
<td>53.155</td>
<td>6.022</td>
<td>51.672</td>
<td>7.316</td>
<td>23.117</td>
<td>46.088</td>
<td>6.979</td>
<td>41.509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price</td>
<td>1.114</td>
<td>0.108</td>
<td>0.94</td>
<td>0.158</td>
<td>1.135</td>
<td>0.084</td>
<td>0.700</td>
<td>0.099</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td>695.935</td>
<td>75.360</td>
<td>505.469</td>
<td>115.324</td>
<td>295.377</td>
<td>109.630</td>
<td>876.180</td>
<td>96.610</td>
<td>533.720</td>
<td>95.951</td>
</tr>
<tr>
<td>Ad GRPs per 100 m inhabitants</td>
<td>102.717</td>
<td>182.703</td>
<td>295.477</td>
<td>248.802</td>
<td>27.246</td>
<td>72.479</td>
<td>103.049</td>
<td>139.493</td>
<td>15.572</td>
<td>61.686</td>
</tr>
</tbody>
</table>

- We formulate our model in general terms (with random component at the market level), so that researchers may use it in future applications with several emerging and mature markets. It is feasible even for our 2-market level analysis because, in the longitudinal HLM, time is nested within the brand, which is nested within the market. Thus, we have not 2 (number of markets) observations but 2 × 6 (number of brands per market) × 69 (number of data periods) = 628 observations for estimation.

- While Hanssens et al. (2010) use an AR(3) model, we estimate an AR(1) model because we need to compare a coefficient and its standard error across brands and countries. We verified that the empirical ordering of brands and countries in stickiness is unchanged whether one uses the AR(1) or the AR(3) model.
First, for communication awareness, 3.870% of the variation can be attributed to market differences, and 51.700% to brand differences (Table 4). Thus, we observe a high residual variance (44.430%) in explaining communication awareness; apparently, factors other than market actions influence whether survey respondents recall having seen marketing communication. Relative price (0.681) and distribution (0.299) have similar effects on communication awareness in each country (Table 8). In contrast, the advertising GRP coefficient is significantly different across country markets as the average brand manages to increase communication awareness with advertising in Brazil (0.009), but not in the U.K. (−0.027). In support of H1a, we thus find that responsiveness of communication awareness to ad GRPs is higher for Brazil than for the U.K.

For brand consideration, 0.367% of its variance is explained by differences between markets, 90.100% by brand differences and the remainder by residual variance (Table 4). Summing up the fixed and the random effects between markets (Table 8), we find that the responsiveness of brand consideration to relative price (−0.231) does not differ significantly between markets, but that Brazil shows a significantly higher responsiveness to distribution (0.312 versus 0.260), while the U.K. shows a significantly higher responsiveness to distribution (0.367 versus 0.309). Thus, we find support for H2a: the responsiveness of brand attitudes (i.e., consideration and liking) to marketing communication is lower in Brazil than in the U.K.

For brand consideration, 0.218% of variance is explained by market level differences, and 74.179% by brand differences. The stickiness of brand consideration is not significantly different for Brazil (0.009) versus the U.K. (0.006). For brand liking, 92.516% of variance is explained by market level differences, and 74.179% by brand differences. The stickiness of brand consideration is not significantly different for Brazil (0.009) versus the U.K. (0.006). For brand liking, 51.700% of variance is explained by market level differences, and 92.516% by brand differences. Thus, we observe a high residual variance (44.430%) in explaining communication awareness; apparently, factors other than market actions influence whether survey respondents recall having seen marketing communication. Relative price (0.681) and distribution (0.299) have similar effects on communication awareness in each country (Table 8). In contrast, the advertising GRP coefficient is significantly different across country markets as the average brand manages to increase communication awareness with advertising in Brazil (0.009), but not in the U.K. (−0.027).

In support of H1b, communication awareness is lower in Brazil than in the U.K. With the exception of consideration, the staying power of mindset metrics is higher in the U.K. than it is in Brazil. For brand consideration, 0.218% of variance is explained by market level differences and 74.179% by brand differences. The stickiness of brand consideration is not significantly different for Brazil (0.009) versus the U.K. (0.006). For brand liking, 92.516% of variance is explained by market level differences and only 4.079% by brand differences. Brand liking has significantly lower stickiness in Brazil (0.184) than in the U.K. (0.759), a result that is contrary to our hypothesis. Gains in brand liking enjoy a multiplier of 2.571 [1/(1 − 0.611)] in Brazil and 8.197 in the U.K.

For communication awareness, 13.869% of the variance is explained by market level differences, and 12.385% by brand differences (Table 4). In support of H1b, communication awareness is lower sticky (Table 8) in Brazil (0.611) than in the U.K. (0.879). Absent new stimuli, gains in communication awareness enjoy a multiplier of 2.571 [1/(1 − 0.611)] in Brazil and 8.197 in the U.K. For brand consideration, 0.218% of variance is explained by market level differences and 74.179% by brand differences. The stickiness of brand consideration is not significantly different for Brazil (0.009) versus the U.K. (0.006). For brand liking, 92.516% of variance is explained by market level differences and only 4.079% by brand differences. Brand liking has significantly lower stickiness in Brazil (0.184) than in the U.K. (0.759), a result that is contrary to our hypothesis. Gains in brand liking enjoy a multiplier of 2.571 [1/(1 − 0.611)] in Brazil and 8.197 in the U.K. Thus, we fail to find support for H2b that brand attitude stickiness is higher in Brazil than in the U.K.

Fig. 3 visualizes the differences in mindset metric stickiness for Brazil versus the U.K. With the exception of consideration, the staying power of mindset metrics is higher in the U.K. than it is in Brazil.

### 4.3. HLM results on sales conversion

The conversion of consumer attitudes into brand sales shows average elasticities (Table 7) of 0.133 for communication awareness (significant at the 5% level), 0.400 for brand consideration and 0.879 for brand liking (both significant at the 1% level). These estimates are lower than, but in the same order as, the average elasticities Srinivasan et al. (2010).
The importance of market variation is striking for the sales conversion equation where 98.574% of the variance is explained by market differences, and 0.447% by brand differences (Table 4). Thus, sales conversion depends on the country, not on the specific brand. The low error (0.979%) in the sales conversion equation implies that a given change in a mindset metric (e.g., 10% increase) has essentially the same sales effect at any time in our data period. We do not report for France: 0.44 for communication awareness, 0.78 for brand consideration and 1.03 for brand liking.

Table 5

### Maximum likelihood estimates of responsiveness equations in longitudinal HLM

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (DV = Log_CA)</th>
<th>Model 2 (DV = Log_Consideration)</th>
<th>Model 3 (DV = Log_Liking)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>-2.366</td>
<td>0.445</td>
<td>-1.989</td>
</tr>
<tr>
<td>Log_Price</td>
<td>0.681</td>
<td>0.108</td>
<td>0.026</td>
</tr>
<tr>
<td>Log_Distribution</td>
<td>0.300</td>
<td>0.067</td>
<td>0.026</td>
</tr>
<tr>
<td>Log_GRPs</td>
<td>-0.009</td>
<td>0.019</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi^{(1)}$</td>
<td>0.054</td>
<td>0.031</td>
<td>0.008</td>
</tr>
<tr>
<td>$\psi^{(2)}$</td>
<td>0.558</td>
<td>0.772</td>
<td>0.163</td>
</tr>
<tr>
<td>$\sqrt{\theta}$</td>
<td>0.522</td>
<td>0.251</td>
<td>0.056</td>
</tr>
<tr>
<td>$\sigma_\alpha^{(Log_Price)}$</td>
<td>0.029</td>
<td>0.008</td>
<td>0.004</td>
</tr>
<tr>
<td>$\sigma_\alpha^{(Log_Distribution)}$</td>
<td>0.008</td>
<td>0.038</td>
<td>0.024</td>
</tr>
<tr>
<td>$\sigma_\alpha^{(Log_GRPs)}$</td>
<td>0.022</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-327.674</td>
<td>-35.282</td>
<td>574.901</td>
</tr>
<tr>
<td>LR test</td>
<td>$\chi^2 = 262.61$, prob &gt; $\chi^2 = 0.000$</td>
<td>$\chi^2 = 711.90$, prob &gt; $\chi^2 = 0.000$</td>
<td>$\chi^2 = 381.15$, prob &gt; $\chi^2 = 0.000$</td>
</tr>
</tbody>
</table>

4.4. Managerial implications

How can managers use the type of elasticities provided in Table 8? The estimated elasticities for each arrow in Fig. 1 may be combined to assess the expected sales gain of contemplated marketing actions in different countries. For example, increasing distribution by 10% increases liking by 0.540% in Brazil (Table 8), but by 1.000% in the U.K. Due to the higher liking stickiness in the U.K., this translates into a long-term liking gain of 0.662% in Brazil and 4.145% in the U.K. Finally, due to the higher sales conversion of liking in the U.K., this long-term liking increase translates into a long-term sales gain of 0.406% in Brazil versus 4.859% in the U.K. Similar calculations for communication awareness and consideration reveal a sales gain of 1.422% and 2.497% in Brazil versus 1.985% and 2.018% in the U.K., respectively. From this analysis, managers learn that the sales impact of distribution occurs mostly through liking in the U.K., but mostly through consideration and communication awareness in Brazil. This knowledge helps them to focus on the most relevant mindset metrics in each market. Moreover, the elasticity magnitudes provide benchmarks for any specific marketing campaign in each market. For instance, it appears ill-advised to criticize a Brazilian manager for failing to increase liking (which is key to sales gain in the U.K.) when, instead, the focus in Brazil should be communication awareness and consideration gains. Thus, our model enables managers to prioritize different metrics in different markets rather than promoting a one-size-fits-all strategy.

5. Discussion and conclusion

This paper presented and illustrated a conceptual framework of how effectiveness criteria for consumer mindset metrics operate differently in an emerging market versus a mature market. Based on regulative, cultural and economic differences between countries, we formalized...
our hypotheses on (1) the responsiveness and stickiness of communication awareness, (2) the responsiveness and stickiness of brand attitudes, and (3) the sales conversion of brand liking. As a first empirical assessment of the framework, we analyze these effects for a major emerging market, Brazil, versus a major mature market, the U.K.

We find support for H1a, H1b, H2a and H3. Brand attitude consideration and liking are less responsive to marketing communication in Brazil than they are in the U.K. Brand liking has a lower sales conversion, while communication awareness has a higher sales conversion in Brazil than in the U.K. Moreover, communication awareness has a higher responsiveness to advertising in Brazil, which is the likely reason for its lower stickiness (Bettman, 1979; Burke & Srull, 1988). However, low responsiveness to advertising does not necessarily imply high stickiness. Contrary to H2b, brand liking is less sticky in Brazil than in the U.K. This may be due to the dynamic demographics in Brazil, which have both a younger population than the U.K. (World Factbook, 2012) and witnesses a rapid increase of the middle class (Broido, Hoevel & Stul, 2012). As a result of such demographics, many consumers in emerging markets are first-time buyers in a product category (Batra, 1999; Burgess & Steenkamp, 2006). Such dynamic demographics may lead to higher instability in the brand liking metric. Future research is needed to examine whether the lower stickiness of liking also applies to other emerging markets.

How would our results hold up in other emerging and mature markets? As our framework is general, we would predict the same responsiveness and conversion differences for emerging (mature) markets low (high) in regulative protection, individualism and income. However, different combinations of these institutional context factors would generate interesting new predictions. For instance, Spain scores relatively low among mature markets on Hofstede’s (1980) scale of individualism (51/100, ranked 20th), while India scores relatively high among emerging markets (48/100, ranked 21st). At the same time, regulative protection and income differ substantially between these two countries, consistent with the U.K.–Brazil difference. Would Spain and India show similar advertising responsiveness of brand attitudes, but different advertising responsiveness of communication awareness and different sales conversion of liking? While our framework would predict so, future research is needed to verify this prediction. The regulative protection and income factors offer the additional benefit of studying changes over time, as the gap between emerging and mature markets is more likely to shrink with respect to those institutional context dimensions than with respect to the cultural factor (Inglehart & Baker, 2000). If consumers in an emerging country receive much better regulative protection against poor-quality products, would they start paying less attention to marketing communication? And if economic prospects substantially deteriorate in a mature market, will the sales conversion of liking decline? These predictions derive from our framework, which requires further empirical validation. Below we discuss important research and managerial implications in the event such validation is successful.

Regarding the advice for brands to become “romantic and mysterious” love marks (Roberts, 2005), our findings imply that the rewards of such a strategy may be much greater in a mature market such as the U.K. than in an emerging market such as Brazil. Indeed, the recent empirical finding that brand liking is highly responsive to advertising and converts strongly into sales (Hanssens et al., 2010) comes from a country (France) where most consumers are high in individualism, income and protection against poor-quality products. Our findings offer reason to believe that the Hanssens et al. (2010) result may not hold for consumers low in individualism and/or income. In our study, price increases liking, but decreases consideration, just as Ferrari may be loved but remain out of reach for many in mature markets, relatively expensive packaged good brands may be liked by emerging market consumers who do not consider buying them in the foreseeable future.

The lower sales conversion of brand liking also implies that a strong emotional connection with consumers may not be as important for

Table 6

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Maximum likelihood estimates of stickiness equation in longitudinal HLM.*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1 (DV=Log_CA)</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Coefficient</td>
</tr>
<tr>
<td>α</td>
<td>−0.122</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.749</td>
</tr>
</tbody>
</table>

* √θ is the standard deviation of the random intercept at the market level, √ρ is the standard deviation of the residuals. ρ(AR(1),market level) is the standard deviation of the slope parameter for the lagged DV at market level, ρ(AR(1),brand level) is the standard deviation of the slope parameter for the lagged DV at brand level.

Random effects

| | Coefficient | SE | z | p>|z| |
|---|---|---|---|---|
| √θ | 0.16 | 0.30 | 0.008 |
| √ρ | 0.13 | 0.42 | 0.071 |
| √σβ | 0.32 | 0.25 | 0.055 |
| ρ(AR(1),market level) | 0.129 | 0.030 | 0.102 |
| ρ(AR(1),brand level) | 0.016 | 0.079 | 0.233 |
| Log likelihood | −133.814 | −30.592 | 595.608 |
| LR test | χ² = 12.73, prob > χ² = 0.013 | χ² = 76.04, prob > χ² = 0.000 | χ² = 71.28, prob > χ² = 0.000 |

Table 7

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Maximum likelihood estimates of sales conversion equation in longitudinal HLM.*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(DV=Log_Sales)</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Coefficient</td>
</tr>
<tr>
<td>α</td>
<td>7.470</td>
</tr>
<tr>
<td>Log_CA</td>
<td>0.133</td>
</tr>
<tr>
<td>Log_Consideration</td>
<td>0.400</td>
</tr>
<tr>
<td>Log_Liking</td>
<td>0.879</td>
</tr>
</tbody>
</table>

* √θ is the standard deviation of the random intercept at the market level, √ρ is the standard deviation of the residuals. ρ(AR(1),brand level) is the standard deviation of the slope parameter for the log of communication awareness, etc.
brands in emerging markets as it is for brands in mature markets (though we acknowledge that, also in mature markets, several brands with a utilitarian focus succeed7). In this context, Western branding experts should exercise care when claiming that “China has no brands in any real sense” (Yong, 2005) and that Chinese consumers are “unable to define the features of a brand” as “the emotional connection is simply absent” (Lindstrom, 2011). Thus, we agree with Cayla and Arnould (2008, p. 7) and question the assumption of prominent marketing practitioners and academics that “the principles of building a strong brand are basically the same across cultures”. Similarly, brands born in emerging markets should be wary of carrying their assumptions into mature markets. For example, Hyundai now recognizes the need in the US market to move beyond a “left-brain choice” (value, fuel economy and lengthy warranty); as a result, it has begun to air ads that “aim to add an emotional connection and remind [people that] buying a Hyundai isn’t just a rational choice” (Ad Age, 2011).

What does this mean for marketing managers? First, patience is gold in an emerging market such as Brazil. That is, managers should immediately track whether consumers received the message, but then they need to give the social influence process time to flourish. Second, a pulse of the GRP spending should allow marketing communication to start the social influence process, which then requires little if any further stimulation due to the effect of word-of-mouth and stickiness in brand consideration. Third, a large portion of the marketing budget should aim to ensure that relevant consumer groups are aware of and consider the brand for purchase.

Our framework and empirical analysis have several limitations that require further investigation. First, we base our propositions on consumer mindset (demand-side) metrics, without explicitly accounting for supply side considerations, such as infrastructure and political stability, or company factors, such as organizational absorption of the marketing concept (e.g., Nakata & Sivakumar, 2001), managerial focus (e.g., Adler & Bartholomew, 1992, Morris & Pitt, 1994; Peterson, 1993) and degree of marketing program standardization (Jain, 1989). Second, our focus on aggregate-level mindset metrics requires us to infer the impact of individual consumer characteristics — controlled experiments are needed to directly demonstrate these links. Third, we use specific operationalizations of regulative, cultural and economic dimensions, and future research may investigate other measures. We do not expect our findings to be sensitive to alternative operationalizations — for one, we obtain similar results substituting Hofstede’s (1980) individualism—collectivism scales with Schwartz’s (1999) autonomy versus relatedness scales. Fourth, our empirical study only considers one product category in one emerging versus one mature market. Further studies are needed to determine the generalizability of our findings across markets and categories. Fifth, we base our empirical analysis on the same mindset metrics for the emerging and the mature market. Different mindset metrics could play a key role in emerging markets. Sixth, as consumer heterogeneity is substantial in both analyzed countries, further research may distinguish among regions, age cohorts and consumer segments. Last but not least, the question remains whether differences in mindset dynamics and marketing effectiveness are mostly driven by institutional, cultural or economic differences. The economic gap between currently emerging and mature markets may disappear within the next few decades, but the cultural differences are likely to remain (Inglehart & Baker, 2000). As countries transition from industrial to service-oriented economies, will consumers continue to choose the ‘safe bets’, i.e., brands with reliable quality and good service (Zhou, 2008) or will they be attracted to the ‘love marks’, i.e., brands that are “romantic, sensual and intimate” (Roberts, 2005)? Time will tell how this romance versus reliability dilemma will evolve.

---

7 We thank the editors for this insight.

---

Table 8
Elasticity estimates (combining fixed and random effects) for Brazil versus U.K.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Brazil</th>
<th>U.K.</th>
<th>Brazil</th>
<th>U.K.</th>
<th>Brazil</th>
<th>U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log_CA</td>
<td>−0.117</td>
<td>−0.117</td>
<td>−0.172</td>
<td>−0.172</td>
<td>−0.094</td>
<td>0.082</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.611</td>
<td>0.878</td>
<td>0.499</td>
<td>0.486</td>
<td>0.184</td>
<td>0.759</td>
</tr>
<tr>
<td>Log_Price</td>
<td>−2.363</td>
<td>−2.369</td>
<td>−1.990</td>
<td>−1.988</td>
<td>−0.295</td>
<td>−0.297</td>
</tr>
<tr>
<td>Log_Distribution</td>
<td>0.681</td>
<td>0.681</td>
<td>−0.231</td>
<td>−0.231</td>
<td>0.127</td>
<td>0.126</td>
</tr>
<tr>
<td>Log_GRPs</td>
<td>0.299</td>
<td>0.299</td>
<td>0.312</td>
<td>0.260</td>
<td>0.054</td>
<td>0.100</td>
</tr>
<tr>
<td>Log_Liking</td>
<td>0.009</td>
<td>−0.027</td>
<td>0.007</td>
<td>0.009</td>
<td>−0.002</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Sales Conversion

<table>
<thead>
<tr>
<th>Metric</th>
<th>Brazil</th>
<th>U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log_Sales</td>
<td>3.836</td>
<td>11.103</td>
</tr>
<tr>
<td>Log_Consideration</td>
<td>0.185</td>
<td>0.081</td>
</tr>
<tr>
<td>Log_Liking</td>
<td>0.613</td>
<td>1.171</td>
</tr>
</tbody>
</table>

---

Fig. 2. Advertising responsiveness of mindset metrics in Brazil versus U.K.

Fig. 3. Stickiness of mindset metrics in Brazil versus U.K.

Fig. 4. Sales conversion of mindset metrics in Brazil versus U.K.
The “green” side of materialism in emerging BRIC and developed markets: The moderating role of global cultural identity

Yuliya Strizhakova a,⁎, Robin A. Coulter b,1

a School of Business, Rutgers University — Camden, 227 Penn Street, Camden, NJ 08102, United States
b School of Business, University of Connecticut, 2100 Hillside Road, Storrs, CT 06269-1041, United States

A R T I C L E   I N F O

Article history:
First received in 30, June 2011 and was under review for 2 ½ months
Available online 16 October 2012

Keywords:
Environmental
Environmentalism
Global identity
Green marketing
Sustainability

A B S T R A C T

Drawing on cultural identity theory, global consumer culture theory, and sustainability research, we examine the “green” side of materialism in emerging BRIC markets and developed (U.S. and Australian) markets. We assess the moderating effect of global cultural identity on the relationship between materialism and environmentally friendly tendencies using three different conceptualizations and measures of global cultural identity — the lifestyle and brand dimensions of global consumption orientation and global connectedness. In emerging markets, we observe strong positive effects of materialism on the concern for environmentally friendly products, the willingness to pay extra for environmentally friendly products, perceptions of global companies as environmentally friendly, and the likelihood to engage in environmentally friendly tendencies for the global segment across all three conceptualizations of global cultural identity; in addition, for individuals with a global cultural identity, we observe a significant positive relationship between materialism and these measures of environmentally friendly tendencies. In developed markets, significant effects are observed only for the global segment, but specific effects depend on the conceptualization of a global cultural identity. Therefore, our results indicate that multinational companies focused on combining materialistic appeals with their green positioning in the emerging markets must carefully target consumers with a strong global cultural identity.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Global sustainability is widely recognized as encompassing the intrinsically interrelated environmental, social, and economic sustainability (Adams, 2006; Lélé, 1991). The interplay between these three pillars of sustainability means that changes to one have downstream effects on the other two. For example, although economic development plays a key role in alleviating world poverty and in building social development, it can come at a cost to the environment (Mabogunje, 2002). As globalization has developed around the world, some have argued that the global integration of national economies further erodes the capacity of individual countries to balance environmental, economic, and social choices and have suggested that the power of sustainable development shifts toward multinational corporations and the global marketplace (Adams, 2006; Paelke, 2005).

In pursuing their global business goals, multinational corporations focus on economic development and thus fuel the growth of materialism worldwide (Belk, Ger, & Askegaard, 2003; Ritzer, 2007; Sharma, 2011; Steenkamp & de Jong, 2010). Notably, the importance of possessions is becoming more evident in emerging markets, but also continues to be apparent in developed countries (Dholakia & Talukdar, 2004; Ger & Belk, 1996; Speck & Roy, 2008). As it relates to sustainability, materialistic consumption fuels the economy but puts a strain on environmental resources and, at the aggregate level, has been deemed to be ecologically destructive (Brown & Kasser, 2005; Kasser, 2005). At the individual level, research documents that a more materialistic individual is less likely to be environmentally conscious (Brown & Kasser, 2005; Good, 2007). Similarly, research examining the relationship between materialism as a value and various manifestations of environmentalism has reported that materialistic individuals engage in fewer environmentally friendly activities (Kilbourne & Pickett, 2008; Richins & Dawson, 1992) and leave a larger ecological footprint (Brown & Kasser, 2005).

Globalization processes, however, also highlight a truth about consumption: it is impossible to consume without limits in an ecologically limited world. As multinational corporations frequently undertake environmentally responsible brand positioning (Osterhus, 1997), individuals around the world have become attentive to ecologically friendly behaviors, and “green” is becoming the new “cool”
across the globe ("Be green, be cool, & be nice to the environment," 2010; "Cool is the new green," 2010). Despite a lesser developed infrastructure to support environmentally friendly behaviors in many emerging (versus developed) markets, there is evidence of an "environmental" segment of more affluent, technologically savvy, and globally oriented consumers interested in "greener" products ("Are emerging market consumers engaging with the green bandwagon?" 2007). Adams (2006) further argues that sustainability (including economic, social, and environmental sustainability) must be understood as a fundamental cultural idea. Therefore, two consumer values linked to sustainability, materialism and environmental consciousness, and their related behaviors, may not necessarily be opposed to each other at the individual level, and their relationship must be further investigated in a global context.

In this paper, we draw upon cultural identity theory, global consumer culture theory, and sustainability research to discuss the relationship between materialism and "environmentally friendly tendencies," defined to include not only market-based tendencies (i.e., concern about environmentally friendly products, willingness to pay more for environmentally friendly products, and perceptions about global brands as environmentally friendly) but also the more general likelihood to engage in environmentally friendly behaviors. We contend that an individual's global cultural identity, that is, the extent to which an individual's identity focus is more global than local (Berry, 2001; Jensen, 2003; Steenkamp, Batra, & Alden, 2003; Strizhakova, Coulter, & Price, 2011), moderates the relationship between materialism and environmentally friendly tendencies. We focus on global cultural identity because individuals more engaged in a global (relative to local) consumer culture are likely to concurrently focus not only on economic growth and the value of their possessions but also on the welfare of the global environment. Specifically, we posit that individuals with a stronger global cultural identity will exhibit a stronger positive relationship between materialism and environmentally friendly tendencies. Furthermore, we speculate that this relationship will be stronger in emerging (than developed) markets where both economic and environmental sustainability emerged simultaneously in response to global integration; we expect a weak relationship in developed markets where these two dimensions of sustainability were historically regarded as mutually exclusive (Adams, 2006; Grove, 2002; Hoffmam, 1997). To test these propositions, we conducted our research using an online panel of adult consumers in the emerging BRIC market and the developed markets of the U.S. and Australia.

Our research integrates theoretical perspectives on sustainability, materialism, and global consumer culture and contributes to a broadened understanding of the "green side" of materialism in several important ways. First, our work situates the relationship between materialism and these environmentally friendly tendencies within a global consumer culture; we argue that materialism and environmentally friendly tendencies can coexist, particularly for a segment of consumers with a global cultural identity who are engaged in global discourses related to both status and product ownership as well as ecologically conscious consumption practices. Second, our examination of the "green side" includes not only the consideration of the likelihood to engage in environmentally friendly behaviors but also three market-based environmental tendencies (i.e., concern for environmentally friendly products when making purchases, the willingness to pay extra for environmentally friendly products, and the perceptions of global companies as environmentally friendly) that have not previously been examined in the context of materialism. Third, we address a recent call to examine consumer behavior in emerging markets that have different socio-historic and cultural developments compared to developed Western markets but are exposed to similar globalization processes and strategies by multinational firms (Burgess & Steenkamp, 2006). We further explore differences in the sizes and composition of cultural identity segments across emerging and developed markets. In the following sections, we elaborate on the theoretical underpinnings of our research, provide additional information about our sample and survey, report our findings, discuss the managerial implications of our research, and identify opportunities for future work.

2. Conceptual framework

As globalization has evolved, cultural identity has been at the core of consumer culture research (Alden, Steenkamp, & Batra, 2006; Kjeldgaard & Askegaard, 2006; Steenkamp & de Jong, 2010; Strizhakova, Coulter, & Price, 2008, 2012; Varman & Belk, 2009; Zhao & Belk, 2008). A key conclusion of this research is that cultural identity should consider an individual's global and local identities (Arnett, 2002) because although consumers may react favorably to symbols of a global consumer culture, they do so in relation to local cultural symbols (Akaka & Alden, 2010; Ger & Belk, 1996; Hung, Li, & Belk, 2007; van Ittersum & Wong, 2010).

Researchers have recently begun to examine an individual's cultural identity, defined as the coexistence of a broad range of beliefs and behaviors embedded to varying degrees in local and global discourses (Strizhakova et al., 2012). Researchers also have offered varying conceptualizations and measurement of this global–local cultural identity, which are based on engagement in global and/or local consumer culture. Some have proposed separate measures of global identity and local identity (Tu, Khare, & Zhang, 2012; Zhang & Khare, 2009). Others suggest that global and local identities can be combined in more intricate ways, such that these identities coexist with a broad range of beliefs, and behaviors embedded in local and global discourses (Strizhakova et al., 2012). Specifically, Steenkamp and his colleagues (Alden et al., 2006; Steenkamp & de Jong, 2010) and Strizhakova et al. (2012) classify individuals as having a cultural identity that is global, glocal, local (national), or alienated (unengaged). The global group includes individuals who prefer a modern global lifestyle, give attention and value to global brands, and have a stronger affiliation with global consumer culture at the expense of local culture. The glocal group includes individuals who are fluent across global and local cultural spaces and who adeptly combine global and local cultural lifestyles and brands. The local group includes individuals who adhere to local traditions and avoid global brands; they have a more nationalistic and ethnocentric orientation. Finally, the alienated group appears to be disinterested in global and local consumer culture and brands. In general, global and glocal consumer segments are more open to the influences of globalization and global consumer culture, whereas local and alienated segments are less open to such influences.

Because materialism and environmentally friendly tendencies are promulgated by multinational firms around the world as reflections of their sustainability mission (encompassing economic, environmental, and social sustainability), we suggest that assessing this relationship in the context of global cultural identity will provide interesting insights. From the perspective of sustainable development (Adams, 2006), individuals who are engaged in the global consumer culture can pursue a more affluent and materialistic lifestyle while also desiring to engage in a "greener" lifestyle, exhibiting more environmentally friendly beliefs and consumption practices. Furthermore, although we expect a personal identity affiliation with the global consumer culture to transcend emerging and developed markets, we expect the relationship to be stronger in emerging than developed markets due to historical differences in the evolution of sustainability in these markets. In pursuing both economic and environmental sustainability, multinational corporations have brought the value of materialistic possessions along with their focus on green products to emerging markets in recent years. Consequently, globally oriented consumers in these emerging markets are more likely to integrate
both materialistic and environmental values that are associated with global brands. Although multinational corporations pursue similar strategies for their global brands in developed markets, it may be more difficult for consumers to integrate economic growth and environmental strategies that have historically clashed at the macro level in these markets. Therefore, we expect global cultural identity to have a stronger effect in emerging markets than in developed markets. Fig. 1 presents our theoretical model, which highlights global cultural identity as an important moderator of the effect of materialistic values on environmentally friendly tendencies in the following sections, we briefly review research on materialism and environmentally friendly tendencies, paying attention to globalization and cultural identity.

2.1. Materialism and global cultural identity

Materialism has been defined as a “set of centrally held beliefs about the importance of possessions in one's life” (Richins & Dawson, 1992, p. 308) or as “the importance a consumer attaches to worldly possessions” (Belk, 1985, p. 291). With consumerism as a focus in the developed markets, materialism has been of interest for three decades. As the emerging markets are increasingly engaged in a global consumer culture, they are witnessing materialism on the rise (Appadurai, 1990; Belk et al., 2003; Ritzer, 2007). Indeed, multinational firms encourage consumerism and promulgate materialism as part of their economic sustainability mission in emerging markets (Dholakia & Talukdar, 2004; Sharma, 2011). Across both developed and emerging markets, materialism has primarily been considered in relation to an individual's general well-being (Rindfleisch, Burroughs, & Wong, 2009); however, other work has examined materialism in the context of consumption beliefs and practices. For example, Rindfleisch et al. (2009) reported that more materialistic individuals have a broader set of self- and communal-brand connections, particularly when they feel insecure and vulnerable. Consumer researchers have also established a positive relationship between materialism and the value-expressive function of foreign products among Chinese consumers (Hung, Gu, & Yim, 2007), between materialism and conspicuous consumption among American consumers (Wang & Wallendorf, 2006), and between materialism and consumer self-enhancement values among Turkish consumers (Karabati & Cemalciar, 2010).

Global consumer culture theory suggests that individuals with a more global focus, that is, a global cultural identity, would be more cosmopolitan and more concerned about how they compare with others around the world and thus claims that the consumption of global brands and living a more cosmopolitan global lifestyle is linked with materialistic tendencies (Alden et al., 2006; Belk, 1985; Hannerz, 2000). As globalization progresses, materialistic individuals embrace a more global and less local cultural identity (Dholakia & Talukdar, 2004; Sharma, 2011). Other work has demonstrated that global and glocal segments have stronger materialistic values and more positive attitudes toward global brands (Alden et al., 2006; Riefler, 2012; Steenkamp & de Jong, 2010) as well as greater acculturation to global consumer culture (Cleveland & Laroche, 2007).

2.2. Environmentally friendly tendencies and global cultural identity

Despite an ongoing tradition in sociology and environmental psychology, the investigation of environmental consciousness and environmentally friendly tendencies has traditionally received less attention in marketing (Ellen, Wiener, & Cobb-Walgren, 1991; Pickett, Kangun, & Grove, 1993). However, more recently, there have been calls for a renewed focus on environmental consumption and sustainability (Kotler, 2011; Prothero et al., 2011). Research has focused on profiling general environmental concerns, attitudes, and behaviors (Dembkowski & Hamner-Lloyd, 1994; Polonsky, 2011), and there is a trend to study consumers’ concern for and consumption of environmentally friendly products (Cornelissen, Dewitte, Warlop, & Vzerbyt, 2007, Cornelissen, Pandraeere, Warlop, & Dewitte, 2008; Goldstein, Galdini, & Griskevicius, 2008; Welsch & Kühling, 2009).

As globalization has evolved, multinational corporations and global brands have emphasized the importance of environmental sustainability in their business mission and green consumerism through their green positioning and advertising appeals (“Are emerging market consumers engaging with the green bandwagon?,” 2007). As such, they further facilitate consumer concern for environmentally friendly products and develop perceptions of global companies and brands as environmentally conscious. Global consumer culture theory

---

**Fig. 1.** Conceptual model: the moderating effect of global cultural identity on the relationship between materialism and environmentally friendly tendencies in emerging and developed markets.
suggests that individuals with a more global focus, that is, a global cultural identity, are more concerned about the environment and more likely to engage in environmentally friendly tendencies. Environmental concern, in addition to freedom, liberty, and human rights, is becoming an important pillar of global citizenship and a more pressing global issue and expression of global culture (Leiserowitz, Kates, & Parris, 2006; Osterhus, 1997; Thompson, 2005; Whalley, 2008). Russell and Russell (2010) refer to both global citizenship and environmental consciousness as superordinate consumer values that are reflective of overarching consumer orientations toward belonging to the global world and the natural environment.

2.3. Rethinking the relationship between materialism and environmentally friendly tendencies

As we have noted, in recent years, firms worldwide have been stimulating both materialistic values and environmental consumption as part of their sustainability mission. On the aggregate level, greater consumption of goods is associated with adverse effects on the environment (Brown & Kasser, 2005; Kasser, 2005), and some researchers have argued that more materialistic individuals are less likely to engage in environmentally friendly tendencies (Kasser, 2002; Kilbourne & Pickett, 2008). We have drawn upon global consumer culture theory (Askegaard, 2006; Jensen, 2003; Steenkamp et al., 2003; Strizhakova et al., 2011) and sustainability perspectives (Adams, 2006) to argue that individuals who exhibit a stronger global cultural identity are more likely to hold materialistic values and be more engaged in environmentally friendly tendencies. In addition, we expect that a global cultural identity moderates the effects of materialism on environmentally friendly tendencies and that the effect is positive and stronger (weaker) among consumers with a stronger (weaker) global cultural identity.

Further, because we expect that global cultural identity is focused on globalization practices related to consumption, we expect this moderating effect to hold in both emerging and developed markets. For many years, economic and environmental sustainability have been perceived as being at odds with one another due to the adverse effect of increased consumption on the environment in developed markets (Grove, 2002; Hoffmam, 1997). Only recently have globalization processes highlighted the interdependence and more complex relationships between economic and environmental sustainability (Adams, 2006). For consumers in developed markets, it may be more difficult to integrate the historical tensions between environmental and economic sustainability and the new interdependent sustainability perspective than for consumers in emerging markets, who have been exposed to the ideas of materialism and environmentalism more recently due to globalization. Therefore, we predict a stronger moderating effect of global cultural identity in emerging markets than in developed markets.

3. Method

3.1. Sample and procedures

The net sample of our study included 1872 adults from the emerging BRIC market (Brazil = 319, Russia = 328, India = 305, China = 295) and the developed market (USA = 302, Australia = 323) who participated in online data collection (approximately 200 participants were removed from the sample due to nonresponse or poor quality responding). Participants had resided for a minimum of seven years in their respective countries, and we targeted an equal number of males and females, as well as equal numbers of participants in each of three age groups: 18–30, 31–45, and 46–60 (we set an upper age limit at 60 because of the lower average lifespan in some markets). Participants had various educational and professional backgrounds (see Table 1 for participant profiles in the emerging BRIC and developed markets and by country), but more educated segments were overrepresented in comparison to the general population in emerging markets due to online data collection. The overall response rate (completed surveys/invitations to participate) across countries was 42%.

We developed the survey questionnaire in English, and the English version was used for our data collections in the U.S., Australia and India, where English is the primary language of schooling. Native speakers translated the questionnaire into Portuguese, Russian, and Mandarin, and then other native speakers back-translated it into English for our data collections in Brazil, Russia and China, respectively. The survey asked a variety of consumption-related questions, including our measures of materialism, global cultural identity, and questions related to environmentally friendly tendencies. To minimize common method biases (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), the software program randomized the order of questions and blocks. Further, to ensure participants’ attention to the survey, we inserted two quality-control questions at the beginning and at the end of the survey. Participants who failed to mark a specific response to these questions were automatically dropped from the study. To further assess response bias, we included two thematically unrelated questions (“Orange is my favorite color,” “I like eating raw vegetables”); correlations between these questions and other questions were non-significant. The marketing research firm eliminated those who provided a uniform response to all questions and also ensured that participants completed the survey only once.

3.2. Measurement

3.2.1. Measurement of latent constructs

Our independent variable, materialism, was measured using six items from the well-established Richins’ scale (1987; see items and scale reliabilities in Appendix A, Table 1A). Our dependent variable, environmentally friendly tendencies, consists of three market-based measures, as well as the likelihood to engage in environmentally friendly behaviors; all were measured on seven-point scales. With regard to the former three measures, we drew upon work related to environmental consumption practices (Cornelissen et al., 2008; Kilbourne et al., 2009; Webb, Mohr, & Harris, 2008). We assessed concern for environmentally friendly products when making purchases with three questions: “When purchasing a new product, how concerned are you that the product is made from: 1) environmentally friendly materials, 2) packaged in biodegradable materials, and 3) made from recycled materials?” The response endpoints were “not at all concerned” and “very concerned.” Willingness to pay extra for environmentally friendly products was assessed with three questions: “How willing are you to pay an extra 3–5% on the top of the price so that the product is made from: 1) environmentally friendly materials, 2) packaged in biodegradable materials, and 3) made from recycled materials?” The response endpoints were “not at all willing” and “very willing.” Perceptions of global companies as environmentally friendly was assessed by the question, “Based on what you know about each of the following six companies, please indicate the extent to which you believe the company is or is not environmentally responsible?” The endpoints were “not at all environmentally responsible” and “very responsible.” Each of the six target brand companies (automotive: Toyota and Mercedes; electronics: Samsung and Apple; food-beverage: Coca-Cola and McDonald’s) has made a statement about its environmental responsibility and sustainability on their respective corporate websites and was identified as one of Interbrand’s top 20 global brands in 2010 (“Top 100 Global Brands,” 2010). Preliminary EFA and CFA analyses confirmed the presence of one underlying construct related to consumer perceptions of global companies as environmentally responsible, with correlated errors between pairs of companies from the same industry (multi-group CFA: $\chi^2(24) = 85.39, \text{CFI} = .98, \text{TLI} = .97, \text{RMSEA} < .05$). Finally, we assessed the likelihood to engage in environmentally friendly behaviors as
related to the following: 1) recycling, 2) conserving use of energy at home, 3) conserving use of water at home, and 4) minimizing household waste/trash to protect the environment. The response endpoints were “not at all likely” and “very likely.”

We used structural equation modeling (AMOS, 17.0) to test the fit of our measurement model composed of five latent constructs (materialism, consumer concern for environmentally friendly products, consumer willingness to pay extra for environmentally friendly products, perceptions of global companies as environmentally friendly, and likelihood to engage in environmentally friendly behaviors). We first tested models individually in each country, and similar to previous research documenting measurement problems with materialism items in cross-cultural settings (Griffin, Babin, & Christensen, 2004; Wong, Rindfleisch, & Burroughs, 2003), we found two items (“It is really true that money can buy happiness” and “People place too much emphasis on material things”) that yielded low loadings (<.50) across all emerging countries. After dropping these two items, the resultant measurement models yielded acceptable fit measurements. Next, we ran a multi-group CFA to establish configural and metric invariance (Steenkamp & Baumgartner, 1998). The fit of the measurement model was acceptable ($\chi^2 = 2619.71$, $df = 93$, $p = .003$). All factor loadings were significant, and all correlations were below .70 (Campbell & Fiske, 1959), indicating configural invariance. Full metric invariance was achieved for all measures ($\Delta \chi^2 = 2916.61$, $df = 93$, $p = .003$). We further confirmed that our latent measures exhibited convergent and discriminant validity (Fornell & Larcker, 1981) (see Appendix A, Table 2A for average variance extracted, internal consistency, and Pearson correlations of our measures).

### 3.2.2. Global cultural identity

We assessed global cultural identity using three measures. The first two reflect an individual’s global lifestyle and brand consumption orientations and have been identified in previous work (Alden et al., 2006; Riefler, 2012; Steenkamp & de Jong, 2010). With regard to these two measures, participants were asked to choose which of four statements (global, glocal, local, and alienated; see Appendix B) most closely represent their personal orientation. The third measure, global connectedness, which focuses on an individual’s overall attachment and belonging to the global world, was developed for this research, based on work related to group and geographic identity (Cameron, 2004; Russell & Russell, 2010). We developed seven seven-point Likert items, tapping into the salience of global world membership and attachment to the global consumer segment (see Appendix B); we then used a median split to categorize participants as having a stronger versus weaker global connectedness.²

To assess the measurement model for each measure, we ran three multi-group SEM analyses for the emerging BRIC market and the developed market data: two across global, glocal, local, and alienated groups for the lifestyle and brand orientations, and one across the strong and weak global connectedness groups. The fit of all measurement models was acceptable for both the emerging BRIC market (lifestyle orientation: $\chi^2 (628) = 1228.15$, CFI = .91, TLI = .90, RMSEA = .04; brand orientation: $\chi^2 (628) = 766.06$, CFI = .91, TLI = .90, RMSEA = .04; global connectedness: $\chi^2 (314) = 554.42$, CFI = .93, TLI = .92, RMSEA = .06) and the developed market (lifestyle orientation: $\chi^2 (628) = 1062.27$, CFI = .94, TLI = .93, RMSEA = .04; brand orientation: $\chi^2 (628) = 1271.55$, CFI = .94, TLI = .93, RMSEA = .04; global connectedness: $\chi^2 (314) = 740.04$, CFI = .96, TLI = .95, RMSEA = .05). Full metric invariance was achieved for all measures (emerging BRIC: lifestyle orientation $\Delta \chi^2 (45) = 91.96$, $p < .01$, brand orientation $\Delta \chi^2 (45) = 87.47$, $p < .01$, global connectedness $\Delta \chi^2 (15) = 16.83$, $p < .05$; developed: lifestyle orientation $\Delta \chi^2 (45) = 59.25$, $p = .05$, brand orientation $\Delta \chi^2 (45) = 66.65$, $p < .01$, global connectedness $\Delta \chi^2 (15) = 23.27$, $p < .01$).

### 4. Results

#### 4.1. Descriptive findings in relation to global cultural identity segments

Table 2 provides profile details of the segments (global, glocal, local, and alienated) for the lifestyle orientation and brand orientation measures of global cultural identity and also for the global connectedness measure, for both the emerging BRIC and developed markets, as well as ANOVA results and post hoc analyses. We note several key findings. First, in relation to segment sizes, we found differences within emerging BRIC and developed markets. With regard to the emerging BRIC market, the glocal segment was largest according to both the lifestyle and brand orientation measures (52.9% and 60.9%, respectively); the other segments were significantly smaller (ranging from 5.9% to 19.2%). In the developed market, the global segment was largest with regard to both the lifestyle and brand orientation measures (59.2% and 60.6%, respectively). In addition, across the emerging BRIC and developed markets, the Chi-squared tests indicate significant variation in segment size with regard to lifestyle orientation ($\chi^2 (4) = 163.87$, $p < .001$) and brand orientation ($\chi^2 (4) = 251.79$, $p < .001$). Table 2 also reports on age, education, international travel, and Internet use (see Footnote 3 in Table 2) among the segments identified with the three measures of global cultural identity. The key differentiating factor across the global cultural identity segments was education; in both emerging BRIC and developed markets, a larger percentage of college-educated participants were in the global and glocal (vs. local and alienated) segments in terms of lifestyle orientation and in the stronger (vs. weaker) global connectedness segment. With regard to brand orientation, in the emerging market, we found that the global and glocal segments had the largest percentage (~80%) of college-educated participants, whereas in the developed market, the global segment had the smallest percentage (~55%) of college-educated participants. Furthermore, participants in the global and glocal segments in the emerging BRIC market reported greater Internet use than the other two segments.
4.2. Assessing the direct effect of materialism on environmentally friendly tendencies

We examined the direct effect of materialism on each of the four environmentally friendly tendencies in the emerging BRIC and developed markets. In both markets, we find a significant positive relationship between materialism and perceptions of global companies as environmentally friendly tendencies (17 and .20, respectively, p < .001), and non-significant relationships with regard to the following: 1) concern for environmentally friendly products, 2) willingness to pay more for environmentally friendly products, and 3) likelihood to engage in environmentally friendly behaviors.

4.3. Testing the moderating effect of global cultural identity

Our model posited a moderating effect of global cultural identity on the relationship between materialism and environmentally friendly tendencies with a stronger effect in emerging markets than in developed markets (Fig. 1). We tested our hypothesized model in AMOS, using the bootstrapping bias-corrected confidence interval procedure (Preacher & Hayes, 2008; Preacher, Rucker, & Hayes, 2007; Zhao, Lynch, & Chen, 2010). The advantage of the bootstrap method is a lack of normality assumption and stronger accuracy of confidence intervals, which are particularly important in smaller samples (Preacher & Hayes, 2008). We used 2000 iterations and set up 95% confidence intervals. We tested three separate models for emerging and developed markets with moderating effects of global cultural identity as expressed through lifestyle orientation (Table 3), brand orientation (Table 4), and global connectedness (Table 5). The fit of the structural model was acceptable for both the emerging BRIC market and the developed market (see Tables 3–5). To compare the moderating effect of global cultural identity on the relationship between materialism and environmentally friendly tendencies within the emerging BRIC market and within the developed market as well as for each individual segment across emerging and developed markets, we ran a series of Chi-square-difference tests (Kline, 1998; see Tables 6 and 7).

Consistent with our expectations, our findings provide evidence that global cultural identity, assessed using three different conceptualizations, has a significant positive impact on the relationship between materialism and environmentally friendly tendencies.

4.3.1. The emerging BRIC market findings

In the emerging BRIC market, we find significant positive effects of materialism on each of three market-based environmentally friendly tendencies (concern for environmentally friendly products, willingness to pay extra for environmentally friendly products, and perceptions of global companies as environmentally friendly), as well as on the likelihood to engage in environmentally friendly behaviors for the global and glocal segments with both lifestyle orientation and brand orientation as moderators and for those with strong global connectedness (see Tables 3–5). Also consistent with our expectations, we observe no significant effects of materialism on the environmentally friendly tendencies for the local and alienated segments or for those with weak global connectedness.

In addition, as shown in Table 6, Chi-squared tests indicate that the effect of materialism with regard to all four environmentally friendly tendencies was significantly stronger for the global segment (vs. the other three segments) using lifestyle orientation as the moderator and for those with stronger (vs. weaker) global connectedness; using brand orientation as the moderator, we observed that the effect of materialism was stronger for the global segment (vs. the other three segments) on two environmentally friendly tendencies (concern for environmentally friendly products and perceptions of global companies as environmentally friendly). In addition, significant differences in the moderating effects of global cultural identity on the relationship between materialism and environmentally friendly tendencies include the following: the glocal (vs. alienated) segment using lifestyle orientation related to the willingness to pay extra for environmentally friendly products; the glocal and local segments (vs. alienated) using brand orientation related to concern for environmentally friendly products and the likelihood to engage in environmentally friendly behaviors.

4.3.2. The developed market findings

In the developed market, the general pattern of results is consistent with our expectations, that is, the relationship between materialism and environmentally friendly tendencies is positive for more versus less globally focused individuals. However, we find some differences depending on the measures of global cultural identity and environmentally friendly tendencies. Specifically, for the global
segment, the effect of materialism on the concern for environmentally friendly products is positive and significant using lifestyle orientation and brand orientation as moderators on the following: willingness to pay extra for environmentally friendly products using brand orientation as the moderator; perceptions of global companies as environmentally friendly using lifestyle orientation and global connectedness as moderators; and likelihood to engage in environmentally friendly behaviors using lifestyle orientation, brand orientation, and global connectedness. In relation to the perception of global companies as environmentally friendly, the effect was stronger in the BRIC (vs. developed) market when brand orientation was the moderator and for those with stronger global connectedness. In relation to the perception of global companies as environmentally friendly, the effect was stronger in the BRIC (vs. developed) market when lifestyle orientation was the moderator. Finally, in relation to the likelihood to engage in environmentally friendly behaviors, the effect of materialism was stronger in the BRIC (vs. developed) market for the glocal segment when lifestyle orientation, brand orientation, and global connectedness were the moderators; and likelihood to engage in environmentally friendly behaviors using lifestyle orientation, brand orientation, and global connectedness was the moderator.

4.3.3. Comparisons across emerging and developed markets within each segment

We predicted stronger moderating effects in the emerging BRIC market than in the developed market (see Chi-squared difference tests reported in Table 7). In relation to the concern for environmentally friendly products and the willingness to pay extra for environmentally friendly products, we observed a positive and significantly stronger effect of materialism in the emerging BRIC (vs. developed) market for the glocal segment when lifestyle orientation was the moderator and for those with stronger global connectedness. In relation to the perception of global companies as environmentally friendly, the effect was stronger in the BRIC (vs. developed) market for the global and glocal segments when lifestyle orientation was the moderator. Finally, in relation to the likelihood to engage in environmentally friendly behaviors, the effect of materialism was stronger in the BRIC (vs. developed) market for the glocal segment with lifestyle orientation as the moderator and for those with stronger global connectedness.

Collectively, our results are clear and consistent in the emerging BRIC market and also provide support for the developed market that individuals with a stronger global cultural identity (assessed via lifestyle orientation, brand orientation, and global connectedness) express a stronger positive relationship between materialism and environmentally friendly tendencies. Stronger effects in the

### Table 3
Lifestyle orientation as the moderator of the materialism and environmentally friendly tendencies relationship: SEM results.

<table>
<thead>
<tr>
<th>Bootstrap bias-corrected method 95%CI</th>
<th>Unstandardized Estimates</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emerging BRIC market</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global segment (n = 142)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>1.03</td>
<td>.20</td>
<td>.73</td>
<td>1.39</td>
<td>.001</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.73</td>
<td>.20</td>
<td>.42</td>
<td>1.07</td>
<td>.001</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.53</td>
<td>.18</td>
<td>.23</td>
<td>.86</td>
<td>.001</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.81</td>
<td>.18</td>
<td>.53</td>
<td>1.13</td>
<td>.001</td>
</tr>
<tr>
<td>Glocal segment (n = 659)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.12</td>
<td>.07</td>
<td>.01</td>
<td>.25</td>
<td>.050</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.19</td>
<td>.08</td>
<td>.05</td>
<td>.33</td>
<td>.016</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.25</td>
<td>.08</td>
<td>.12</td>
<td>.38</td>
<td>.003</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.14</td>
<td>.07</td>
<td>.02</td>
<td>.25</td>
<td>.050</td>
</tr>
<tr>
<td>Local segment (n = 239)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.08</td>
<td>.16</td>
<td>-.19</td>
<td>.33</td>
<td>.660</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.13</td>
<td>.16</td>
<td>-.14</td>
<td>.39</td>
<td>.470</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.19</td>
<td>.13</td>
<td>-.04</td>
<td>.38</td>
<td>.174</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.12</td>
<td>.11</td>
<td>-.07</td>
<td>.29</td>
<td>.276</td>
</tr>
<tr>
<td>Alienated segment (n = 205)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.04</td>
<td>.18</td>
<td>-.18</td>
<td>.31</td>
<td>.859</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>-.16</td>
<td>.16</td>
<td>-.42</td>
<td>.10</td>
<td>.298</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.08</td>
<td>.13</td>
<td>-.13</td>
<td>.30</td>
<td>.554</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.10</td>
<td>.13</td>
<td>-.13</td>
<td>.31</td>
<td>.434</td>
</tr>
<tr>
<td><strong>Developed market</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global segment (n = 37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.84</td>
<td>.21</td>
<td>.36</td>
<td>1.35</td>
<td>.009</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.09</td>
<td>.54</td>
<td>-.75</td>
<td>.98</td>
<td>.823</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.88</td>
<td>.31</td>
<td>.56</td>
<td>1.25</td>
<td>.001</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.71</td>
<td>.24</td>
<td>.35</td>
<td>1.13</td>
<td>.003</td>
</tr>
<tr>
<td>Glocal segment (n = 175)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.01</td>
<td>.19</td>
<td>-.31</td>
<td>.18</td>
<td>.925</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.01</td>
<td>.22</td>
<td>-.35</td>
<td>.37</td>
<td>.970</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.02</td>
<td>.17</td>
<td>-.12</td>
<td>.18</td>
<td>.748</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>-.08</td>
<td>.13</td>
<td>-.29</td>
<td>.14</td>
<td>.570</td>
</tr>
<tr>
<td>Local segment (n = 259)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>-.01</td>
<td>.13</td>
<td>-.22</td>
<td>.21</td>
<td>.958</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>-.07</td>
<td>.17</td>
<td>-.34</td>
<td>.21</td>
<td>.709</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.11</td>
<td>.08</td>
<td>-.02</td>
<td>.25</td>
<td>.163</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.04</td>
<td>.07</td>
<td>-.08</td>
<td>.16</td>
<td>.601</td>
</tr>
<tr>
<td>Alienated segment (n = 154)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>-.39</td>
<td>.13</td>
<td>-.73</td>
<td>-.03</td>
<td>.081</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>-.39</td>
<td>.26</td>
<td>-.82</td>
<td>.03</td>
<td>.123</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.12</td>
<td>.13</td>
<td>-.08</td>
<td>.33</td>
<td>.349</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>-.08</td>
<td>.17</td>
<td>-.37</td>
<td>.19</td>
<td>.628</td>
</tr>
</tbody>
</table>

Model fit: \( \chi^2(628) = 1228.15, CFI = .91, TLI = .90, RMSEA = .04 \)
emerging market than in the developed market are observed for the
glocal segment and for those with stronger global connectedness.

5. Discussion: global cultural identity and the “green side” of materialism

In the era of global sustainability, multinational corporations have focused their business strategies on both economic and environmental sustainability. Although economic development inevitably depletes natural resources, the two pillars of sustainability coexist in more complex ways at the individual level. On the one hand, multinational corporations promulgating economic growth have also been disseminating materialistic values through their glamour and status appeals, particularly in emerging markets. On the other hand, environmental sustainability has become a pressing concern in the current global context, and documenting an inverse relationship between materialism and environmentally friendly tendencies at the individual level by examining the moderating effect of global cultural identity on this relationship in the global context, specifically contrasting the effects in the emerging BRIC market with those in the developed markets of the U.S. and Australia.

As individuals around the world have increasingly been exposed to the global media and global discourse, Arnett (2002), Hermans and Dimaggio (2007), and others have discussed cultural identity formation and the interplay between globalization and local identity construction. In this study, we have built upon prior work on global cultural identity and the interplay between globalization and local identity formation, and others have discussed cultural identity formation and the interplay between globalization and local identity construction.

Table 4

Brand orientation as the moderator of the materialism and environmentally friendly tendencies relationship: SEM results.

| Model fit: $\chi^2$ (628) = 766.06, CFI = .91, TLI = .90, RMSEA < .04 |
|---|---|---|---|---|
| Unstandardized estimates | SE | Lower | Upper |
| Materialism on concern for environmentally friendly products | .58 | .13 | .37 | .81 | .001 |
| Materialism on willingness to pay extra for environmentally friendly products | .18 | .12 | .01 | .37 | .046 |
| Materialism on perceptions of global companies as environmentally friendly | .58 | .15 | .34 | .84 | .001 |
| Materialism on likelihood to engage in environmentally friendly behaviors | .34 | .11 | .18 | .53 | .002 |

Developed market

Global segment (n = 29)

Materialism on concern for environmentally friendly products | .80 | .89 | .09 | 3.32 | .008 |
Materialism on willingness to pay extra for environmentally friendly products | .81 | 1.47 | .18 | 2.76 | .051 |
Materialism on perceptions of global companies as environmentally friendly | .03 | .30 | .17 | .45 | .643 |
Materialism on likelihood to engage in environmentally friendly behaviors | .15 | .61 | .29 | 1.50 | .304 |

Glocal segment (n = 240)

Materialism on concern for environmentally friendly products | -.21 | .15 | .45 | .05 | .189 |
Materialism on willingness to pay extra for environmentally friendly products | -.20 | .20 | .57 | .08 | .263 |
Materialism on perceptions of global companies as environmentally friendly | .15 | .10 | .00 | .33 | .096 |
Materialism on likelihood to engage in environmentally friendly behaviors | .00 | .11 | .18 | .19 | .986 |

Local segment (n = 142)

Materialism on concern for environmentally friendly products | .14 | .19 | .15 | .45 | .400 |
Materialism on willingness to pay extra for environmentally friendly products | .04 | .21 | .34 | .34 | .879 |
Materialism on perceptions of global companies as environmentally friendly | .05 | .09 | .10 | .21 | .587 |
Materialism on likelihood to engage in environmentally friendly behaviors | .13 | .14 | .12 | .33 | .358 |

Alienated segment (n = 214)

Materialism on concern for environmentally friendly products | .13 | .20 | .21 | .44 | .519 |
Materialism on willingness to pay extra for environmentally friendly products | -.13 | .25 | .56 | .22 | .551 |
Materialism on perceptions of global companies as environmentally friendly | .30 | .12 | .13 | .52 | .239 |
Materialism on likelihood to engage in environmentally friendly behaviors | .00 | .10 | .14 | .16 | .971 |

Bootstrap bias-corrected method

95% CI

relationship between materialism and environmental tendencies at the individual level by examining the moderating effect of global cultural identity on this relationship in the global context, specifically contrasting the effects in the emerging BRIC market with those in the developed markets of the U.S. and Australia.

As individuals around the world have increasingly been exposed to the global media and global discourse, Arnett (2002), Hermans and Dimaggio (2007), and others have discussed cultural identity formation and the interplay between globalization and local identity construction. In this study, we have built upon prior work on global consumer culture (Akaka & Alden, 2010; Alden, Steenkamp, & Batra, 1999, Alden et al., 2006; Ger & Belk, 1996; Steenkamp & de Jong, 2010) to propose that global cultural identity is an important moderator of the effects of materialism on environmentally friendly tendencies. In our work, we use three measures of global cultural identity – lifestyle and brand dimensions of global cultural orientation and global connectedness – and show a similar pattern of results across our three measures of global cultural identity. In contrast to some past research documenting an inverse relationship between materialistic and environmental values and practices (Brown & Kasser, 2005; Kilbourne & Pickett, 2008; Richins & Dawson, 1992), our work supports the sustainability perspective and documents a positive relationship between...
Global cultural identity segment comparisons for environmentally friendly tendencies in emerging BRIC market and developed market.

Table 6
Global connectedness as the moderator of the materialism and environmentally friendly tendencies relationship: SEM results.

<table>
<thead>
<tr>
<th>Bootstrap bias-corrected method 95% CI</th>
<th>Unstandardized Estimates</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emerging market</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong global connectedness (n = 608)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.39</td>
<td>.11</td>
<td>.22</td>
<td>.56</td>
<td>.001</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.35</td>
<td>.13</td>
<td>.17</td>
<td>.60</td>
<td>.002</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.37</td>
<td>.10</td>
<td>.19</td>
<td>.53</td>
<td>.003</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.20</td>
<td>.07</td>
<td>.08</td>
<td>.32</td>
<td>.001</td>
</tr>
<tr>
<td>Weak global connectedness (n = 593)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.08</td>
<td>.07</td>
<td>.05</td>
<td>.21</td>
<td>.346</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.06</td>
<td>.08</td>
<td>.06</td>
<td>.19</td>
<td>.423</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.22</td>
<td>.08</td>
<td>.08</td>
<td>.35</td>
<td>.329</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.24</td>
<td>.07</td>
<td>.11</td>
<td>.36</td>
<td>.437</td>
</tr>
<tr>
<td>Model fit: $\chi^2 (628) = 554.42, \text{CFI} = .93, \text{TLI} = .92, \text{RMSEA} = .06$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Developed market</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong global connectedness (n = 302)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>.01</td>
<td>.12</td>
<td>.22</td>
<td>.20</td>
<td>.979</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>-.11</td>
<td>.15</td>
<td>.40</td>
<td>.12</td>
<td>.407</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.21</td>
<td>.08</td>
<td>.09</td>
<td>.34</td>
<td>.010</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>-.06</td>
<td>.10</td>
<td>.21</td>
<td>.09</td>
<td>.555</td>
</tr>
<tr>
<td>Weak global connectedness (n = 302)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materialism on concern for environmentally friendly products</td>
<td>-.04</td>
<td>.15</td>
<td>.28</td>
<td>.19</td>
<td>.786</td>
</tr>
<tr>
<td>Materialism on willingness to pay extra for environmentally friendly products</td>
<td>.04</td>
<td>.18</td>
<td>-.24</td>
<td>.33</td>
<td>.782</td>
</tr>
<tr>
<td>Materialism on perceptions of global companies as environmentally friendly</td>
<td>.15</td>
<td>.09</td>
<td>.01</td>
<td>.30</td>
<td>.076</td>
</tr>
<tr>
<td>Materialism on likelihood to engage in environmentally friendly behaviors</td>
<td>.05</td>
<td>.09</td>
<td>-.08</td>
<td>.18</td>
<td>.540</td>
</tr>
<tr>
<td>Model fit: $\chi^2 (314) = 740.04, \text{CFI} = .96, \text{TLI} = .95, \text{RMSEA} = .05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Materialism and perceptions of global brands as environmentally friendly and non-significant relationships between materialism and three other environmentally friendly tendencies (i.e., concern for environmentally friendly products, consumer willingness to pay extra for environmentally friendly products, and likelihood to engage in environmentally friendly behaviors).

Consistent with our expectations, we find the “green side” of materialism in individuals with a strong global cultural identity in both the emerging BRIC and developed markets, but the effects are stronger in emerging markets. In emerging markets, individuals with a global cultural identity are engaged by possessions and also show concern for environmentally friendly products. This global segment has perceptions of global companies as being more environmentally friendly, is willing to pay extra for environmentally friendly products, and is more likely to engage in environmentally friendly behaviors, such as recycling and the conservation of natural resources. These effects are strong and consistent across all three measures of global cultural identity. Additionally, in the emerging BRIC market (but not the developed market), we find the “green side” of materialism among individuals with a strong glocal cultural identity. We speculate that individuals in the emerging market glocal segment may have a heightened attention to globalization relative to those in the

Table 6
Global cultural identity segment comparisons for environmentally friendly tendencies in emerging BRIC market and developed market.

<table>
<thead>
<tr>
<th>$\chi^2$ – difference test for environmentally friendly tendencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td><strong>Emerging BRIC market</strong></td>
</tr>
<tr>
<td>Global vs. glocal*</td>
</tr>
<tr>
<td>Global vs. local</td>
</tr>
<tr>
<td>Global vs. alienated</td>
</tr>
<tr>
<td>Glocal vs. local</td>
</tr>
<tr>
<td>Glocal vs. alienated</td>
</tr>
<tr>
<td>Local vs. alienated</td>
</tr>
<tr>
<td><strong>Developed market</strong></td>
</tr>
<tr>
<td>Global vs. glocal</td>
</tr>
<tr>
<td>Global vs. local</td>
</tr>
<tr>
<td>Global vs. alienated</td>
</tr>
<tr>
<td>Glocal vs. local</td>
</tr>
<tr>
<td>Glocal vs. alienated</td>
</tr>
<tr>
<td>Local vs. alienated</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Global connectedness (strong vs. weak)</strong>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging BRIC market</td>
</tr>
<tr>
<td>Developed market</td>
</tr>
</tbody>
</table>

* The reported $\Delta \chi^2$ tests in each row indicate whether there is a significant difference between the two identified segments on each of the environmentally friendly tendencies.

* $p < .05$.

** $p < .01$.

*** $p < .001$. 

Comparison of environmentally friendly tendencies for global cultural identity segments for the emerging BRIC market versus the developed market.

Table 7

<table>
<thead>
<tr>
<th>Global cultural identity segment</th>
<th>Concern for environmentally friendly products</th>
<th>Willingness to pay extra for environmentally friendly products</th>
<th>Perceptions of global companies as environmentally friendly</th>
<th>Likelihood to engage in environmentally friendly behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifestyle orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>.02</td>
<td>.67</td>
<td>7.05**</td>
<td>.19</td>
</tr>
<tr>
<td>Glocal</td>
<td>.60</td>
<td>1.15</td>
<td>6.60**</td>
<td>4.57**</td>
</tr>
<tr>
<td>Local</td>
<td>.10</td>
<td>.59</td>
<td>.24</td>
<td>.41</td>
</tr>
<tr>
<td>Alienated</td>
<td>2.53</td>
<td>.66</td>
<td>.02</td>
<td>.55</td>
</tr>
<tr>
<td>Brand orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>.35</td>
<td>2.87</td>
<td>1.43</td>
<td>.54</td>
</tr>
<tr>
<td>Glocal</td>
<td>5.95*</td>
<td>3.73*</td>
<td>.05</td>
<td>1.90</td>
</tr>
<tr>
<td>Local</td>
<td>.64</td>
<td>3.29</td>
<td>.03</td>
<td>2.90</td>
</tr>
<tr>
<td>Alienated</td>
<td>2.99</td>
<td>.00</td>
<td>2.16</td>
<td>.01</td>
</tr>
<tr>
<td>Global connectedness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong</td>
<td>6.06**</td>
<td>4.87*</td>
<td>2.83</td>
<td>6.54**</td>
</tr>
<tr>
<td>Weak</td>
<td>.25</td>
<td>1.21</td>
<td>.01</td>
<td>2.24</td>
</tr>
</tbody>
</table>

*Significant effects were stronger in emerging BRIC than developed market.
* p < .05.
** p < .01.

In the developed market, we find support for the positive effect of materialism on all environmentally friendly tendencies for the global segment; however, the significance of this effect on each of the four dependent variables varies across the different conceptualizations of global cultural identity. Perhaps it is more difficult for global consumers in the developed market to negotiate economic and environmental sustainability, having a longer history of being focused on the macro-level adverse effects of consumption on the environment. In addition, we note that the global segments in the developed market are much smaller than those in the emerging market and that larger samples may yield a more consistent pattern of results. As expected in both the emerging BRIC and developed markets, we do not find the “green side” of materialism for the local and alienated segments or for those with weak global connectedness.

6. Managerial implications

In pursuing their economic and environmental sustainability goals, multinational firms face significant challenges in managing their brands and corporate communications around the world. Despite the economic advantages of pursuing a global citizenship strategy, our results indicate that multinational corporations must carefully contemplate their messages in local markets, attending to country-level designations (e.g., emerging vs. developed) and individual-level variables that are important in the context of global cultural identity (Menon & Menon, 1997; Osterhus, 1997; Thompson, 2005). More specifically, our findings document that the importance of possessions and environmentalism resonate with those consumers who identify with the global consumer culture and are less tied to local influences, particularly in emerging markets. These consumers easily integrate economic and environmental sustainability appeals, and multinational firms would do well to communicate messages about status, consumption, and environmental themes to this segment. However, in developed markets, the global segment represented the smallest percentage of respondents (4.6% to 6%), but this is perhaps not surprising, given the relatively young history of global culture and a strong local consumption history in developed markets.

The glocal segment in the emerging BRIC market (accounting for a majority [53% to 61%] of study participants) exhibited a pattern similar to that of the global segment but with weaker effects. Because of its size, the glocal segment is an important target for multinational firms, but attention must be given to the fact this segment is attuned to values of both global and local cultures. To better target this segment, multinational companies would benefit from a more comprehensive assessment of global and local values and the extent to which global versus local culture influences consumption. The global and glocal segments in the emerging markets have attained a college education, enjoy travel, and engage with the Internet. Based on our findings, multinational firms would benefit from promoting an image of being environmentally friendly, engaging these segments in ecologically conscious consumption practices, and considering differing pricing models because of their willingness to pay extra for environmentally friendly products.

Our results indicate non-significant relationships between materialism and environmentally friendly tendencies for the local and alienated segments and for those with weak global connectedness in both the developed and emerging BRIC markets. These segments comprise a critical mass of study participants in the developed market and emerging BRIC market. Consequently, multinational firms must carefully contemplate the promotion of their economic and environmental sustainability to these audiences, as they appear to resist globalization influences.

7. Future research directions

Our work draws attention to several avenues of future research. First, our research supports sustainability research (Adams, 2006) that predicts the coexistence of economic and environmental sustainability. Although materialism and environmentally friendly tendencies appear to be incompatible at the macro level, individuals who are globally oriented and more open to messages by multinational corporations exhibit both materialistic and environmentally friendly tendencies. For example, an investigation of whether globally oriented materialistic consumers truly care about the environment or are simply responding to a fashionable “green” trend is of interest. Furthermore, our results indicate that the global and glocal segments in the emerging market are young and upwardly mobile with access to global media (through the Internet) and global culture (by traveling abroad). Additional investigations related to this cohort, constructs related to materialism, such as conspicuous consumption, and the importance of status and fashion trends in identity creation would be of interest. On a related note, understanding more about how firms can manage the delicate balance between materialistic values and consumption...
within the context of economic sustainability and environmentalism is a worthy undertaking.

Second, we conceptualized global cultural identity as the extent to which an individual's identity focus is more global than local, and we assessed our model using three measures—lifestyle and brand dimensions of consumption orientation and global connectedness. As we noted, global cultural identity has been conceptualized and measured in a variety of ways, and there continue to be discussions regarding the best approach to measuring global cultural identity (Alden et al., 1999, 2006; Steenkamp et al., 2003; Strizhakova et al., 2012; Zhang & Khare, 2009). Because we observed some differences in moderating effects across our different measures of global cultural identity in the developed market, future research may explore the boundary conditions of our model using alternative measures of global cultural identity.

Third, there is a well-documented gap between environmentalism as a value and what people say versus what they actually do in terms of their environmental actions and behaviors (Alwitt & Pitts, 1996). Our work considered market-based environmental tendencies related to concerns about and the willingness to pay for environmentally friendly goods and perceptions of global companies promoting environmentally responsible agendas, as well as the likelihood to engage in environmentally friendly behaviors such as recycling and resource conservation. To more carefully assess socially desirable responses, an interesting extension of our research would be to track the development of ecologically friendly infrastructure in an emerging market and use longitudinal surveys and diaries to concurrently track an individual's global cultural identity as well as environmentally friendly attitudes and behaviors. Additionally, the use of experiments to understand how individuals with differing cultural identities react to global versus local environmental concerns and their willingness to pay for products that are associated with environmental causes would provide additional information useful to firms producing environmentally friendly products and organizations promoting ecologically conscious behavior.

Finally, we used online panels to collect our data, which lead to more educated segments being overrepresented in comparison to the general population in the emerging markets. Furthermore, we requested quota sampling on age and gender to match the samples on these variables. Future research using a representative sample from emerging and developed markets would enable firms to better understand the size of the global, glocal, local, and alienated segments in markets of interest. This information would certainly be valuable to firms contemplating targeting specific audiences with regard to the promotion of consumerism and environmental agendas.

8. Conclusion

Sustainability is receiving increased attention around the globe, and as multinational corporations pursue global citizenship positioning strategies, understanding the role of global cultural identity around the world, particularly within emerging markets, is critical. Our research calls into question the perspective that materialism and environmentally friendly tendencies are incompatible and support current sustainability research by documenting the role of global cultural identity as a moderator of the relationship at the individual level. We found that individuals with a strong global cultural identity in both developed and emerging markets and those with a strong glocal cultural identity in emerging markets concurrently displayed materialism and environmentally friendly tendencies. As global and local firms engage in developing the emerging marketplace, understanding the evolving nature of cultural identity, global consumer culture, and three dimensions of sustainability will be of growing interest.

Appendix A

Table 1A

| Metrically-invariant measurement model: factor loadings, reliabilities, and model fit. |
|----------------------------------|----------------------------------|----------------------------------|
| Unstandardized factor loadings   | Emerging BRIC market | Developed market |
| Brazil                       | Russia                        | India                          | China                          | US                    | Australia                     |
| Materialism                   | 1.00                          | 1.00                          | 1.00                          | 1.00                          | 1.00                          | 1.00                          |

Environmental friendly tendencies

Concern for environmentally friendly products

When purchasing a new product, how concerned are you that the product is:

- Made from environmentally friendly materials.
- Packaged in biodegradable materials.
- Made from recycled materials.

Reliability: .70 .71 .70 .75 .80 .75 .75

Willingness to pay extra for environmentally friendly products

Suppose you are buying a product that typically costs $5. How willing are you to pay extra 3–5% on top of the price so that the products is:

- Made from environmentally friendly materials.
- Packaged in biodegradable materials.
- Made from recycled materials.

Reliability: .92 .70 .79 .92 .95 .79 .87

Perceptions of global companies as environmentally friendly

Based on what you know about each of the following companies, please indicate the extent to which you believe the company is or is not environmentally responsible:

- Mercedes-Benz
- Toyota
- Coca-Cola
- McDonald's
- Samsung
- Apple, Inc.

Reliability: .88 .84 .83 .88 .86 .89 .89

Likelihood to engage in environmentally friendly behaviors

In general, how likely are you to:

- Recycle
- Conserve your use of energy at home
- Conserve your use of water at home
- Minimize household waste/trash to protect the environment.

Reliability: .85 .83 .79 .86 .88 .87 .87

χ² (df) = 2619.71 (942)

CFI/TLI = .93/91

RMSEA = .03

Means are not presented for individual countries because scalar invariance was not achieved.

a Six items from the Richins’ (1987) materialism scale were included in the survey; the four items reported here were used in the final analyses. Two other items ("It is really true that money can buy happiness" and "People place too much emphasis on material things") were dropped because of cross-cultural measurement problems.

b Prices were modified across countries to reflect local prices in local currency for an average price for a 1 liter bottle of Coke. Euromonitor Global Market Information Database was used to derive the average price for each country.
Table 2A

<table>
<thead>
<tr>
<th></th>
<th>CR</th>
<th>AVE</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging BRIC market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.76</td>
<td>.45</td>
<td>.19 (.04)</td>
<td>.14 (.02)</td>
<td>.18 (.03)</td>
<td>.27 (.07)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.91</td>
<td>.77</td>
<td>.51 (.26)</td>
<td>.43 (.18)</td>
<td>.49 (.24)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.91</td>
<td>.77</td>
<td>.51 (.26)</td>
<td>.43 (.18)</td>
<td>.49 (.24)</td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.87</td>
<td>.54</td>
<td>.24 (.08)</td>
<td>.17 (.03)</td>
<td>.13 (.02)</td>
<td>.11 (.01)</td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.88</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.72</td>
<td>.40</td>
<td>.13 (.02)</td>
<td>.17 (.03)</td>
<td>.13 (.02)</td>
<td>.11 (.01)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.72</td>
<td>.47</td>
<td>.55 (.30)</td>
<td>.23 (.05)</td>
<td>.52 (.27)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.76</td>
<td>.52</td>
<td>.29 (.08)</td>
<td>.39 (.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.83</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.86</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.76</td>
<td>.45</td>
<td>.24 (.08)</td>
<td>.12 (.01)</td>
<td>.17 (.03)</td>
<td>.26 (.07)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.87</td>
<td>.67</td>
<td>.58 (.34)</td>
<td>.26 (.07)</td>
<td>.41 (.17)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.88</td>
<td>.71</td>
<td>.20 (.04)</td>
<td>.29 (.08)</td>
<td>.09 (.01)</td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.84</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.81</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.76</td>
<td>.45</td>
<td>.20 (.04)</td>
<td>.14 (.02)</td>
<td>.44 (.19)</td>
<td>.32 (.10)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.80</td>
<td>.57</td>
<td>.65 (.42)</td>
<td>.43 (.18)</td>
<td>.52 (.27)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.82</td>
<td>.59</td>
<td>.46 (.21)</td>
<td>.43 (.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.83</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.83</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developed market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.80</td>
<td>.50</td>
<td>.03 (.00)</td>
<td>.04 (.00)</td>
<td>.20 (.04)</td>
<td>.00 (.00)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.96</td>
<td>.88</td>
<td>.66 (.43)</td>
<td>.26 (.07)</td>
<td>.62 (.38)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.97</td>
<td>.92</td>
<td>.18 (.03)</td>
<td>.47 (.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.87</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.88</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Materialism</td>
<td>.77</td>
<td>.46</td>
<td>.00 (.00)</td>
<td>−.09 (.01)</td>
<td>.09 (.01)</td>
<td>.11 (.01)</td>
</tr>
<tr>
<td>2. Concern for environmentally friendly products</td>
<td>.95</td>
<td>.87</td>
<td>.58 (.34)</td>
<td>−.05 (.00)</td>
<td>.53 (.28)</td>
<td></td>
</tr>
<tr>
<td>3. Willingness to pay extra for environmentally friendly products</td>
<td>.97</td>
<td>.95</td>
<td>.00 (.00)</td>
<td>.40 (.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perceptions of global companies as environmentally friendly</td>
<td>.88</td>
<td>.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Likelihood to engage in environmentally friendly behaviors</td>
<td>.88</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix B. Assessment of global cultural identity

1. Measure of the lifestyle orientation of global cultural identity (Alden et al., 2006). Participants were asked to choose one of the four statements that best represents their personal orientation. The statements were as follows: 1) global: “It is important for me to have a lifestyle that I think is similar to the lifestyle of consumers in many countries around the world rather than one that is more unique to or traditional in (my country),” 2) glocal: “I try to blend a lifestyle that is considered unique to or traditional in (my country) with one that I think is similar to the lifestyle of consumers in many countries around the world,” 3) local: “It is more important for me to have a lifestyle that is unique to or traditional in (my country),” and 4) alienated: “I couldn’t care less about the countries associated with any brand; brand names mean nothing to me.”

2. Measure of the brand orientation of global cultural identity (Steenkamp & de Jong, 2010). Participants were asked to choose one of four statements that best represents their personal orientation. The statements were as follows: 1) global: “I prefer to buy both local brands that are sold only in (my country) and brands that I think are bought by consumers in many countries around the world,” and 2) glocal: “I prefer to buy both local brands that are sold only in (my country) and brands that I think are bought by consumers in many countries around the world,” and 3) local: “I prefer to buy local brands that are sold only in (my country) rather than brands that I think are bought by consumers in many countries around the world,” and 4) alienated: “To be honest, I do not find the typical lifestyle in (my country) or the lifestyles of consumers in other countries very interesting.”

3. Measure of global connectedness. Global connectedness which taps into the salience of global group membership and consumer attachment to the global group was measured by developing items from previous research on group and geographic identity (Cameron, 2004; Russell & Russell, 2010). Participants were asked to express their agreement with the following statements on a seven-point scale.

<table>
<thead>
<tr>
<th>Global Connectedness</th>
<th>Unstandardized factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging</td>
<td>Developed</td>
</tr>
<tr>
<td>I have a strong attachment to the global world.</td>
<td>Marker</td>
</tr>
<tr>
<td>I feel connected to the global world.</td>
<td>.99</td>
</tr>
<tr>
<td>I think of myself as a global citizen.</td>
<td>1.12</td>
</tr>
<tr>
<td>It is important to me to feel a part of the global world.</td>
<td>1.13</td>
</tr>
<tr>
<td>Thinking about my identity, I view myself as a global citizen.</td>
<td>1.10</td>
</tr>
<tr>
<td>Feeling like a citizen of the world is important to me.</td>
<td>1.18</td>
</tr>
<tr>
<td>I would describe myself as a global citizen.</td>
<td>1.18</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>.94</td>
</tr>
<tr>
<td>( \chi^2/df )</td>
<td>170.43/40</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.06</td>
</tr>
</tbody>
</table>


Conducting field research in subsistence markets, with an application to market orientation in the context of Ethiopian pastoralists

Paul T.M. Ingenbleek a,⁎, Workneh Kassa Tessema b, Hans C.M. van Trijp a

a Wageningen University, The Netherlands
b Haramaya University, Ethiopia

Article history:
First received in 13, September 2011 and was under review for 7 1/2 months
Available online 23 December 2012

Keywords:
Research methodology Bias Equivalence Subsistence markets Emerging markets Market orientation

1. Introduction

In the past decade, marketing research has witnessed a growing interest in emerging markets (e.g., Erdem, Swait, & Valenzuela, 2006; Johnson & Tellis, 2008; Kwak, Jaju, & Larsen, 2006; Steenkamp & Burgess, 2002). Researchers have also begun to pay particular attention to subsistence markets within emerging markets: that is, to the four billion people living in subsistence conditions at the base of the income pyramid (Burgess & Steenkamp, 2006; Prahalad & Hammod, 2002) who account for more than 40% of the gross national income of emerging markets (Schneider, 2004). In addition to their resource scarcity, consumers and (micro-)entrepreneurs in subsistence markets often lack formal institutional support, such that they have developed their own informal networks (Viswanathan, Rosa, & Ruth, 2010). This institutional distance from high-income markets (e.g., Rivera-Santos & Rufin, 2010; Van den Wajenberg & Hens, 2012) challenges the generalizability of marketing theories that have emerged from high-income economies (Burgess & Steenkamp, 2006) as well as the validity and usefulness of extant research methods that provide a basis for most theory testing (e.g., Gau, Jae, & Viswanathan, 2012; Viswanathan, Gau, & Chaturvedi, 2008).

A few pioneers have conducted marketing research in subsistence settings (Arnould, 1989; Shultz & Shapiro, 2012; Van Tilburg, 2010), and this research stream has received new impetus from initiatives concerning transformative consumer research (e.g., Mick, Pettigrew, Pechmann, & Ozanne, 2012), international corporate social responsibility research (e.g., Smith, Bhattacharya, Vogel, & Levine, 2010), and consumer literacy investigations (Viswanathan, 2012). Such initiatives have prompted further qualitative research exploring some of the significant contextual differences between high-income and subsistence markets (e.g., Abdelnour & Branzei, 2010; Arnould, 2001; Arnould & Mohr, 2005; Kambewa, Ingenbleek, & Van Tilburg, 2008; Viswanathan, Rosa, et al., 2010). In turn, the collective qualitative insights from these studies have pushed research on subsistence markets to a more mature stage that supports quantitative research approaches, theory testing, and quantified insights into the mechanisms of subsistence markets. However, some authors recommend taking certain measures when conducting market research in subsistence markets (Krämer & Beltz, 2008; Sridharan & Viswanathan, 2008; Viswanathan, Sridharan, & Ritchie, 2008), such as accounting for context-specific measurement issues (e.g., Guesalaga & Marshall, 2008; Toledo-López, Díaz-Picardo, Jiménez-Castañeda, & Sánchez-Medina, 2012) and adapting techniques for experimental studies to the institutional context (Gau et al., 2012; Hounhouigan, Ingenbleek, Van der Lans, Van Trijp, & Linemann, 2012; Viswanathan, Gau, et al., 2008).

Burgess and Steenkamp (2006) outline the basic scientific processes that are inherent to studying marketing theories in emerging markets: namely, theory development, the acquisition of meaningful data, data analysis, and learning. We focus on expanding knowledge about the data acquisition and analysis steps by noting the challenges that arise from field research in subsistence markets that potentially impact equivalence and bias. Moreover, we provide guidance on how to address these challenges. Through this approach, we aim to extend the methodological marketing literature to subsistence markets (e.g., Churchill,
Table 1
Institutional influences of subsistence markets on the research process.

<table>
<thead>
<tr>
<th>Institutional subsystem</th>
<th>Emerging markets (Burgess &amp; Steenkamp, 2006)</th>
<th>Subsistence markets</th>
<th>Potential biases</th>
<th>Research considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socioeconomic subsystem</strong></td>
<td>Diversity</td>
<td>Extreme differences in household size and income, living standards, access to human development resources.</td>
<td>Resource scarcity is relatively common.</td>
<td>Construct</td>
</tr>
<tr>
<td></td>
<td>Method (sample)</td>
<td>Remote respondents may be overlooked and are more difficult to reach due to weak infrastructure.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method (sample)</td>
<td>Resource scarcity may put a stigma on respondents that makes them try to avoid research participation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method (administration)</td>
<td>Compensation of respondents in terms of financial rewards or social capital is often necessary.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demographics</td>
<td>Young, growing, large pool of undereducated people.</td>
<td>Subsistence populations are the main source of the young, growing, and large pools of under-educated people.</td>
<td>Method (sample)</td>
</tr>
<tr>
<td></td>
<td>Method (administration)</td>
<td>Low education may affect the quality of interviewers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamics</td>
<td>Rapid social, political, and economic change.</td>
<td>Mobility (in particular urbanization).</td>
<td>Method (sample)</td>
</tr>
<tr>
<td></td>
<td>Method (administration)</td>
<td>Tracing people that are mobile and that may change names is difficult, in particular in longitudinal studies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method (administration)</td>
<td>Interviewing people in different physical environments (noise, temperature) may cause bias.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regulative subsystem</strong></td>
<td>Rule of law</td>
<td>Moderate abuse of office for private gain, reliance on legal rights enforceable in courts of law (including investor rights) are lower, and legal outcomes are more unlikely.</td>
<td>Informal rules and laws may contradict and fill the institutional gaps of formal ones.</td>
<td>Method (sample)</td>
</tr>
<tr>
<td></td>
<td>Stakeholder influence on corporate governance</td>
<td>Government, civil society, supply chain stakeholders influence high.</td>
<td>NGOs and government extensions may remove structural holes.</td>
<td>Method (sample)</td>
</tr>
<tr>
<td></td>
<td>Method (sample)</td>
<td>When the social capital balance is negative for the nonprofit organization, specific groups of respondents may drop out.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cultural subsystem</td>
<td>Hierarchy versus egalitarianism</td>
<td>Hierarchy emphasized.</td>
<td>Being acculturated means coping with uncertainty, hierarchy is relatively high.</td>
</tr>
<tr>
<td></td>
<td>Embeddedness versus autonomy</td>
<td>Embeddedness emphasized</td>
<td>Embeddedness is an acculturated means to cope with resource</td>
<td>Method (sample)</td>
</tr>
</tbody>
</table>
To illustrate typical challenges that arise when performing fieldwork in subsistence markets and to suggest appropriate responses, we present a study on the market orientation–performance relationship among the subsistence population of pastoralists in East Africa (i.e., Ethiopia). The relationship between market orientation and performance is among the most studied in the marketing literature (Cano, Carrillat, & Jaramillo, 2004; Kirca, Jayachandran, & Bearden, 2005). However, because subsistence contexts differ dramatically from the North American and Western European contexts in which market orientation theory was developed, the market orientation and performance constructs and the rationale for their relationships may differ as well. Our empirical application illustrates typical methodological challenges arising in subsistence marketplaces and suggests means of overcoming them. We show how the generalizability of marketing theories can be assessed in this challenging context, which differs remarkably from high-income Western markets in which most marketing theories have been developed.

In the next sections, we detail the institutional characteristics of subsistence markets and how they affect field research through their impact on bias and equivalence. After we consider how the characteristics of subsistence markets might influence the market orientation–performance relationship, we offer several hypotheses. These hypotheses are then formally tested in an empirical study. We conclude with a discussion and several implications, including concrete suggestions for further quantitative studies in subsistence contexts.

2. The institutional environment of subsistence markets

Burgess and Steenkamp (2006) identify typical institutional subsystems in emerging markets and compare these to institutional subsystems of high-income markets. We build on their work by outlining typical institutional subsystems (socioeconomic, regulative, and cultural) in subsistence markets (see Table 1). For example, in their socioeconomic institutional environment, subsistence markets feature significant resource scarcity: that is, limited access to key production factors, such as infrastructure, information, capital, and knowledge (e.g., Jimenez, 1995; Sheth, 2011; World Bank, 2005). This scarcity is amplified by high population growth due to high fertility and decreasing mortality rates, which places increasing pressure on resources, including the educational system (Bloom, Canning, & Sevilla, 2003). In their search for resources, actors in subsistence markets become relatively mobile in space and time. For example, many individuals are employed in multiple locations by relying on flexible work hours or seasonal employment (Yach, Mathews, & Buch, 1990). The most significant evidence of such mobility is migration from rural areas, where people depend in particular on subsistence agriculture, to urban areas (Montgomery, 2008). However, resources in urban areas are becoming similarly scarce because economic growth rarely can keep pace with the exponential growth of urban populations (Cohen, 2006). Furthermore, migrants to urban and peri-urban areas usually face the challenge of adapting to the demands of life in densely populated towns and settlements, where the culture and lifestyle are often very different. Migrants often form subcultures based on their original locations and maintain ties with their place of origin as a “fall-back” option (e.g., Gugler, 2002; Lesetedi, 2003).

A regulative subsystem establishes formal rules to maintain stability, order, and continuity (Burgess & Steenkamp, 2006). However, governments and legislation exert limited influence over subsistence markets (Castells & Portes, 1989), a situation that creates severe institutional gaps (Chen, 2007; Rivera-Santos, Rufin, & Kolk, 2012). These gaps, in turn, hinder the enforcement of formal rules, the support for economic activities, the governance of transactions, and the protection of people’s rights to use scarce natural resources, such as agricultural land and water (e.g., Bigsten, Kimuyu, & Lundvall, 2004; Hitt, Dacin, Levitas, Arregle, & Borza, 2000). Because these regulative institutional gaps also prevent the development of supporting industries, such as finance or distribution (Rivera-Santos & Rufin, 2010), marketing systems in subsistence markets typically exhibit structural holes or voids (see, for example, Abdelnour & Branzei, 2010). Because of their experience in developing and maintaining network relationships with actors in subsistence markets (Sheth, 2011; Van den Weyenberg & Hens, 2012), nonprofit governmental and nongovernmental organizations (NGOs) often try to fill these gaps (Rivera-Santos et al., 2012; Sheth, 2011).

Finally, subsistence markets exhibit relatively strong cultural institutions. Although some of these institutions contradict formal regulative institutions (Arnould & Mohr, 2005), they may help to fill regulative gaps (De Soto, 2000), for example, through hierarchy relations and embeddedness (Burgess & Steenkamp, 2006). Hierarchy refers to a cultural emphasis on obeying role obligations within a proscribed and legitimately unequal system of distribution of power, roles, and resources. These obligations help to guarantee responsible behaviors that preserve the social fabric (Licht, Goldschmidt, & Schwartz, 2005; Schwartz, 2006). According to Licht et al. (2005), cultural hierarchy also grants power to authorities who are charged with controlling the uncertainty that is typically associated with resource scarcity and regulative gaps in subsistence markets. Embeddedness refers to a desirable relationship between an individual and a group that helps to maintain the status quo and limits actions that might disrupt group solidarity or a traditional order (Licht et al., 2005; Schwartz, 2006). Various studies have shown that subsistence markets worldwide are characterized by social interactions and network relationships (Chikweche & Fletcher, 2010; Espinoza, 1999; Lesetedi, 2003; Viswanathan, Rosa, et al., 2010). Thus, through embeddedness in traditional ties, including kinship, age groups, and religions, members come to view themselves as collective group entities. They aim to fulfill group goals, cope with resource scarcity to fulfill basic subsistence needs for the group, and achieve a shared way of life (e.g., London & Hart, 2004; Losby, Else, Kingslow, Edgcomb, Malm, & Kao, 2002; Viswanathan, Rosa, et al., 2010; Woolcock & Narayan, 2000). Despite commonalities such as resource scarcity, the weak influence of formal rules and legislation, and high cultural embeddedness and hierarchy, subsistence populations

<table>
<thead>
<tr>
<th>Institutional subsystem</th>
<th>Emerging markets (Burgess &amp; Steenkamp, 2006)</th>
<th>Subsistence markets</th>
<th>Potential biases</th>
<th>Research considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>scarcity and regulative institutional gaps.</td>
<td>Method (sample) With multi-group memberships, stratified sampling may be more difficult. Snowballing is more likely to lead to sampling biases. Method (instrument) Network relationships may be the cause of self-serving biases, in particular in snowball samples. Item Meaning and sensitivity of items may vary between groups. Translation of items requires insight in the cultural backgrounds of potential respondents.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 (continued)
can differ remarkably in the symbols, signs, and behavioral expressions of a culture (cf., Guglielmino, Viganotti, Hewlett, & Cavalli-Sforza, 1995).

Together, these defining characteristics of subsistence markets have considerable impact on equivalence and bias in quantitative field studies. They also have consequences for the generalizability of market orientation theory, as we explain in the following sections.

3. Bias and equivalence in field research in subsistence markets

When investigating the generalizability of marketing theories to subsistence markets, researchers quickly encounter the question of whether the outcomes of their measures of key concepts in subsistence markets allow comparisons with those obtained in other (high-income or subsistence) contexts. This is a question of equivalence (Van de Vijver & Leung, 2011). Bias undermines the equivalence of measurement outcomes across research contexts, and only in the absence of bias can constructs and theories be meaningfully compared across research contexts. Not surprisingly, many insights into the types and sources of bias originate in the field of cross-cultural psychology. We build on this research tradition described by Van de Vijver and Leung (2011) but extend beyond it by identifying specific sources of bias that emerge from the typical institutional characteristics of subsistence markets (see Table 1). We offer a priori means of preventing bias during the research design and data collection stages of the research process, and we suggest a posteriori means of identifying and, in some cases, correcting for bias during the analysis stage. We structure our discussion along the lines of the different types of bias distinguished by Van de Vijver and Leung (2011): construct bias, method bias (including sample bias, instrument bias, and administration bias), and item bias.

3.1. Construct bias

Construct bias implies that the construct measured is not identical across research contexts. It may arise when definitions or components of the construct differ between contexts or when the behaviors indicative of a construct differ. Construct bias may arise when resource-rich contexts are compared to resource-scarce contexts because, in the latter context, behaviors are developed to cope with resource scarcity. As these behaviors may differ between groups, construct bias can also occur when researchers make comparisons between subsistence markets. The type of bias and the remedy depend on the objective of the study, among other factors. Studies that aim to make a comparison only at the theoretical level should verify whether the meaning of and the relationships between the constructs are likely to exist in the context in which the theory is tested. In addition to practicing theoretical reflection, researchers may identify such theoretical construct bias a priori through a qualitative prestudy (e.g., Viswanathan, Rosa, et al., 2010). Qualitative prestudy methods include desk research, observation techniques, long interviews and focus groups with subjects, and interviews with experts in local culture and language. Researchers can also use the prestudy to sample behaviors indicative of the construct in the specific context (so-called emic items) to measure the construct and subsequently test whether the hypothesized relationships between them exist.

Studies may also take a comparative approach, by which researchers aim to empirically compare the relative levels of a construct or the strength of a relationship between constructs across contexts. In such cases, researchers should also use their prestudy to assess the potential overlap in behaviors indicative of the construct across groups. Based on their assessment, researchers may adapt measurement instruments so that items that describe general behaviors and that hold construct relevance for all of the study's contexts are sampled (so-called etic items). For example, the World Bank (2012) has developed numerous living standard measures to assess the education, health, employment, and assets owned by individuals, validated across a range of markets. Researchers can also check a posteriori for construct bias either by utilizing multi-group confirmatory factor analysis (CFA) or by comparing the factor solutions from exploratory factor analyses across groups (Steenkamp & Baumgartner, 1998; Van de Vijver & Leung, 2011). However, the more the contexts of a comparison differ, the more difficult it is to identify a set of standardized (etic) items that have the same meaning and relevance across all research contexts, and there may be a need to add or even substitute context-specific emic items.

De Jong, Steenkamp, and Veldkamp (2009) recently introduced a model based on item response theory (IRT) that permits adding emic items to a limited set of etic items while still allowing for meaningful comparisons of construct scores across research contexts.

3.2. Method bias

Method bias may arise when the study is implemented in a subsistence market; this type of bias emerges from the sample, administration, or instrument of analysis.

3.2.1. Sample method bias

Sample method bias arises from the incomparability of samples due to sample characteristics that have a bearing on target measures. As a result, sample differences offer a rival explanation for differences in measurement scores. This type of bias can emerge even when both samples are representative (Van de Vijver & Leung, 2011). For example, when comparing a high-income country to a subsistence market, observed differences in representative samples can be caused by population differences with respect to demographics. In such cases, stratified samples based on age groups may produce more valid comparisons. Other sources of sample bias arise from (1) sampling methods, (2) the selection of key respondents, and (3) several types of non-response bias.

First, the use of simple random sampling as a strategy to ensure sample comparability is virtually impossible in subsistence markets. With their weak regulatory subsystem, subsistence markets typically lack formal institutions such as a Records Office or Chamber of Commerce. As a result, formal sampling frames with information about the size of the research population or the contact details of potential respondents are non-existent (Elahi, 2008). Alternatively, snowball sampling is relatively easy in subsistence markets because respondents have rich network relationships; however, applying this method demands considerable care. When researchers enter a subsistence market to collect data, they should be aware that they become part of the transaction network in which buyers and sellers build social capital by doing favors, such as providing credit (e.g., Viswanathan, Rosa, et al., 2010). Social capital theory explains this phenomenon by suggesting that social networks tie economic value in the form of information and the repayment of social obligations (Coleman, 1988; Herreros, 2004). Social resource theory further argues that networks may embed reputational value—for example, status—which, in turn, provides access to economic resources, such as credit (Khaire, 2010; Linn, 1999).

From a social capital perspective, recommending someone for a paid interview might reflect a return on a favor. Thus, respondents might not recommend the most suitable contacts for the study but rather people to whom they owe a favor, which creates a sampling bias. Snowball sampling is therefore recommended as a secondary sampling method, at best, within pre-defined strata or clusters.

Stratified sampling is particularly applicable in rural subsistence contexts because identifying the different strata and their relative sizes is often easy. Stratified sampling is also important because it may prevent long-distance bias by ensuring that remote communities with unique economic, social, and cultural traditions are not overlooked in the sampling strategy (Koenig & Shepherd, 2001). In urban and peri-urban areas, however, the embeddedness that characterizes subsistence markets renders stratified sampling more complex. According to social identity complexity theory, people in complex societies, such as urban areas, may share an in-group
membership on one dimension (e.g., region of origin) but belong to different categories on another (e.g., economic sector) [Brever & Chen, 2007; Miller, Brewer, & Arbuckle, 2009]. Researchers may, therefore, encounter groups with diverse cultural practices, social networks, and languages in one city, one section of a city, or even a single street (Yach et al., 1990). If identifying strata is not feasible, researchers can apply cluster sampling by first identifying the clusters (e.g., districts, streets) and then randomly drawing a sample of different clusters (Levy & Lemeshow, 2008). Modern information sources such as aerial photography (Google Earth) or geocoding of sampling points using GPS coordinates (e.g., Jacquez & Rommel, 2009) may help researchers to determine relative population sizes and strata or clusters.

Sample method bias may also emerge from the failure to identify the key respondent associated with the appropriate unit of observation. In subsistence markets, individuals may have varied and embedded group memberships (e.g., nuclear families embedded within extended families, clans, and tribes). As decisions may often be made at the group level rather than at the individual level, it can be difficult to discern who is responsible for purchase or sales decisions. Measurement outcomes become inequivalent when scores are not provided by the appropriate respondents, as when decision authority differs between research contexts. As such bias cannot be addressed a posteriori, researchers should determine the appropriate unit of observation or key respondent on the basis of qualitative research prior to their quantitative data collection (e.g., Yach, 1992).

Even when the appropriate respondents have been identified and approached, researchers should be aware that subsistence markets have specific challenges with respect to non-response bias (which leads to biased samples and, thus, a lack of equivalence). First, entering networks of subsistence actors is difficult without an appropriate introduction. Nonprofit organizations can facilitate such introductions on the basis of trust that they have generated from a positive “balance sheet” (Viswanathan, Gau, et al., 2008). However, this approach can also lead to response bias if potential respondents refuse participation because the partner organizations have a negative “balance sheet” with respect to social capital. Toledo, De la Paz Hernández, and Griffin (2010) find, for example, that subsistence entrepreneurs regard governmental support projects with suspicion. Second, given that many subsistence markets are characterized by cultural hierarchies, researchers also may need to seek approval for their study from local authorities, such as chiefs, tribal leaders, or urban block committees. Without such approval, respondents may be more inclined to refuse participation. Third, resource scarcity and a lack of education can stigmatize respondents; thus, respondents may avoid gatherings that could expose their poverty or literacy levels to avoid being labeled as poor or illiterate (Viswanathan, Gau, et al., 2008). Fourth, mobility, which is typical of subsistence markets, may cause response bias in longitudinal studies (Iman & Nuwayhid, 2004; Maluccio, 2004; Thomas, Frankenfeld, & Smith, 2001). Tracing respondents who have moved may be difficult, especially if the respondents adopt different names when they begin to identify with different groups, a phenomenon that is not uncommon in subsistence contexts (Kroeger, 1983). Because the search for resources is rarely random (Maluccio, 2004), researchers should check for response bias by comparing important characteristics of respondents who remain in the sample with the traits of those who drop out. Yang, Zhao, and Dhar (2010) recently proposed a posteriori methods to address underreporting bias in longitudinal studies.

First, as professional fieldwork organizations are typically absent from subsistence markets, the research quality largely depends on the quality and expertise of the interviewers who are specifically recruited for the study and who cannot always be closely monitored during data collection. Potential sources of bias include interviewers’ lack of education, experience and expertise in interviewing, and communicative abilities; moreover, poor instructions and interviewer training may cause bias. Researchers should carefully select their interviewers and use education as a criterion. Extensive training of interviewers with adequate examples and exercise is required to ensure that studies and interviews are administered in the same way across groups. Detailed manuals and administration protocols may need to be developed that specify the research and interview procedures in addition to contingency plans for when problems arise. Detailed instructions and monitoring where possible may help to further reduce administration bias. Finally, researchers may collect data on relevant background variables, including interviewers’ educational level, labels to indicate which interviewer completed each questionnaire, and the interviews’ duration, to allow for a posteriori analysis of administration method bias.

Second, during face-to-face interviews, communication problems may arise from language, vocabulary, and cultural differences. If it is not possible to hire interviewers with a compatible cultural background, it is important to make the interviewer aware of cultural considerations, such as taboo topics and potential hierarchy effects (respondents may adapt their answers to the hierarchical level of the interviewer). In some cultures, hierarchy effects extend to gender differences (Yach, 1992). As communication dynamics are difficult to monitor and control, it is important to record interviewer information with each interview to allow for a posteriori analysis.

A third source of administration method bias may arise from the fact that, in subsistence markets, it is often necessary to (financially) compensate respondents for their participation. Compensating informants can lead to a social response bias, such that respondents provide only the information they believe the interviewer wants to hear, out of gratitude for their remuneration. Even if respondents do not receive financial rewards for their participation, social bias may arise due to social capital. For example, if access to respondents is facilitated through nonprofit organizations that have achieved a positive balance of social capital with respondents due to the services and development aid they provide, a social response bias may occur. Further, such social capital may cause self-serving bias because respondents answer strategically, anticipating that their responses may be forwarded throughout their network (Viswanathan, Sridharan, et al., 2010). In general, researchers need to ensure trust-based relationships with their respondents, clearly explain the purpose of the study and the affiliations of the researcher, assure confidentiality, and confirm that respondents understand that there are no right or wrong answers. They may also vary the rewards paid to respondents according to socioeconomic class, with higher rewards paid to the higher classes (Trujillo, Barrios, Camacho, & Rosa, 2010). Furthermore, they can apply some of the established techniques developed to prevent, detect, and correct biases (see De Jong, Pieters, & Fox, 2010, for a recent contribution).

Fourth, bias can arise from a variation in the physical and social contexts of administration. Variation of the physical environment includes, for example, noise and temperature. The remedy is to conduct the interviews only in relaxed atmospheres, which can be a challenge in studies that focus on subsistence entrepreneurs, who work long days in busy and noisy marketplaces. The social context refers to the presence of other people during the interview. Due to the high level of embeddedness, people may see the interview as an experience to be shared with community members, making it more difficult to obtain respondents’ personal opinions, values, or thoughts. The same limitation may hold when researchers require multiple respondents to minimize single method bias (Rindfleisch et al., 2008; Van Bruggen, Lilien, & Kacker, 2002). This administrative bias indicates the need for respondents to be interviewed in a trusting research environment, particularly when respondents should be
separated from others (Viswanathan, Gau, et al., 2008). However, the presence of others can be advantageous when researchers question respondents about decisions that they usually make as a group (for example, important purchases). In such cases, researchers should be sure to interview respondents in the appropriate context (for example, in the house rather than in public places when the household is the relevant group). If context differences are inevitable, they should be coded during the interview to allow for a posteriori testing.

3.2.3. Instrument bias

*Instrument bias* arises from instrument characteristics such as familiarity with stimulus material, response modes, and response procedures (Van de Vijver & Leung, 2011). A lower level of education and experience among respondents can result in various forms of instrument bias. First, less literate respondents often have difficulty in understanding abstractions (Viswanathan, Gau, et al., 2008; Viswanathan, Rosa, & Harris, 2005) and may perform better on relatively simple tasks that utilize easy-to-understand language (e.g., Viswanathan, Gau, et al., 2008). Schwartz, LeMahieu, Lehmann, Burgess, Harris, and Owens (2001) report, for example, that a shorter and more concrete scale for an abstract construct, such as values, works better in subsistence contexts. Less literate respondents also tend to engage in pictographic thinking (Gau et al., 2012); as a result, experiments and conjoint studies conducted in subsistence contexts will likely benefit from including pictographic stimuli. For example, Chaturvedi, Chiu, and Viswanathan (2009) gave respondents a categorization task with different drawn objects, and in a conjoint study of pineapple attributes, Hounhouigan et al. (2012) developed picture frames for the different attribute levels. Additionally, respondents may have less sophisticated numerical abilities. Research in numerical cognition indicates that two core number systems have emerged through

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Context</th>
<th>Method</th>
<th>Relevant results on market orientation–performance relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhuian (1997)</td>
<td>Saudi-Arabia</td>
<td>Several branches of nine different banks located in major metropolitan areas</td>
<td>Personal interviews</td>
<td>No significant effects</td>
</tr>
<tr>
<td>Burgess and Nyajeka (2007)</td>
<td>Zimbabwe</td>
<td>Harare retailers in the apparel, automotive parts, electronics, footwear, furniture, sporting goods, and supermarket sectors</td>
<td>Personal interviews</td>
<td>Positive main effect.</td>
</tr>
<tr>
<td>Ellis (2005)</td>
<td>China</td>
<td>Exporters of locally made manufactured goods.</td>
<td>Personal interviews</td>
<td>No significant effects.</td>
</tr>
<tr>
<td>Sin et al. (2000)</td>
<td>China</td>
<td>Companies in Beijing</td>
<td>Survey questionnaire</td>
<td>Positive main effects for sales growth and customer retention. No significant effects for return on investment, and market share.</td>
</tr>
<tr>
<td>Singh (2003)</td>
<td>India</td>
<td>Firms with a formal marketing department in New Delhi, Bombay, and Calcutta</td>
<td>Personal interviews</td>
<td>Positive main effects for return on investment and customer retention. No significant effect for foreign market presence. Positive moderating effects of competitive intensity and market dynamism.</td>
</tr>
<tr>
<td>Subramanian and Gopalakrishna (2001)</td>
<td>India</td>
<td>Domestic and multinational manufacturing and service firms</td>
<td>Survey questionnaire</td>
<td>Positive main effects for growth in revenue, success in new products, and ability to retain customers.</td>
</tr>
<tr>
<td>Tse, Sin, Yau, Lee, and Chow (2003)</td>
<td>China</td>
<td>Business companies in Hong Kong with operations in both mainland China and Hong Kong</td>
<td>Survey questionnaire</td>
<td>Positive main effects of customer orientation, competitor orientation, and interfunctional coordination.</td>
</tr>
</tbody>
</table>
Item bias (also referred to as differential item functioning) refers to anomalies at the item level and occurs when two respondents with the same standing on a construct (e.g., they are equally market-oriented) do not have the same mean score on an item because the item has different psychological meanings in different research contexts. Item bias can arise from poor item translation, ambiguities in the original item, low familiarity or appropriateness of the item content in subsistence markets, or connotations associated with the item wording. With the high level of heterogeneity in subsistence markets, the meaning and sensitivity of specific items and questions likely depend on local cultures and traditions (Anker, 1983; Bulmer & Warwick, 1983).

Although it is always advisable to prevent item bias through a priori procedures such as prestudy, proper translation by someone with insight into respondents’ cultural background, and pretesting instruments, advanced methodologies have been developed to identify and analyze item bias and improve the psychometric properties. Studies that include a single group because they make comparisons at the theoretical level only may utilize a single-group CFA model (Steenkamp & Van Trijp, 1991). In addition to the often-used covariance-based model, the use of partial least squares (PLS) approaches can be advantageous, particularly when studies include formative scales and/or when sample sizes are small (see Hair, Sarstedt, Ringle, & Mena, 2012, for a review). For studies that directly compare across research contexts, multi-group CFA is the dominant approach to exploring across-group measurement invariance (see Steenkamp & Baumgartner, 1998). However, IRT-based approaches, such as those developed by De Jong et al. (2008, 2009) and De Jong, Steenkamp, and Fox (2007), provide a more powerful alternative, as they are more sensitive to the identification of item bias and relax some of the assumptions that underlie CFA. Specifically, the IRT model does not require that at least two items exhibit invariance across all research contexts to allow direct comparisons of construct means and construct relations. Moreover, IRT does not build on strong distributional assumptions, and it recognizes the ordinal nature of many commonly applied rating scales.

4. Illustration with market orientation

The positive market orientation–performance relationship is most likely one of the most substantiated relationships in mainstream marketing theory (Cano et al., 2004; Kirca et al., 2005), but our understanding of the full generalizability of the market orientation–performance relationship remains constrained by the lack of testing in subsistence markets. In Table 2, we provide an overview of market orientation studies in emerging markets, which reveals that evidence from Asia and Africa is not conclusive: these studies report positive, negative, and insignificant effects. In emerging markets, studies also tend to focus on large and medium-sized firms embedded in a network of formal institutions located in and around cities with relatively well-developed infrastructures.

To test the market orientation–performance relationship, we adapt the theoretical framework to the characteristics of subsistence populations (see Fig. 1). Building on Narver and Slater’s (1990)
conceptualization of market orientation, we follow contemporary studies (e.g., Frambach, Prabhu, & Verhallen, 2003; Gatignon & Xuereb, 1997; Han, Kim, & Srivastava, 1998; Voss & Voss, 2000) that formulate unique hypotheses for each of the three market orientation components (customer orientation, competitor orientation, and interfunctional coordination). For our subsistence market context, we include a potential moderator: population density. In subsistence populations, rural areas create sparsely populated networks in which actors are distant from one another and have little contact. In contrast, urban areas represent densely populated networks with a high volume of communication among actors (e.g., Podolny & Baron, 1997).

Two characteristics of subsistence populations—regulative institutional and resource scarcity—likely weaken the effect of customer orientation. Without formal regulatory institutions, actors may not be able to appropriate the value they create for customers because they lack protections from malpractices or opportunism (Hitt et al., 2000). Furthermore, resources are needed to create value, but if they are not readily available, the critical success factor in the market may shift from combining resources to simply obtaining resources (Pfeffer & Salancik, 1978). Despite these challenges to the relationship, Licht et al. (2005) argue that embeddedness encourages actors to comply with the unwritten laws of fair business relationships even in the absence of strong regulatory institutions. In addition, customer-oriented actors might see opportunities where others do not. Even in conditions of resource scarcity, actors who are more customer oriented may be able to create more value from the available resources than their non-customer-oriented counterparts (Day & Nedungadi, 1994). As Sheth (2011, p. 173) states, “If necessity is the mother of invention, then resource shortage is the father of innovation.” We hypothesize that despite the challenges of subsistence markets, customer orientation has a positive effect on performance.

**Hypothesis 1.** Customer orientation positively influences performance in subsistence markets.

For competitor orientation, we consider the possibility of a construct bias in subsistence markets because embeddedness emphasizes mutual solidarity and reciprocity rather than rivalry (Schwartz, 1992). For example, a competitor orientation perspective suggests that the members of the Gurung tribe in rural Nepal should compete on markets for food-stuffs and other goods; however, in practice, neighbors, friends, and relatives continuously provide assistance to one another. Refusing requests for help is unthinkable because the tribe considers such tactics harmful to security, belonging, and nurturance (McHugh, 1989). The concept of market competition is likely meaningless in such a culture. However, competitor orientation may be meaningful in another way. Kohli and Jaworski (1990) distinguish between the sense (information acquisition concerning competitors) and response (responding to competitors’ actions) functions of competitor orientations. In some subsistence markets, the sense function remains relevant because the networks in which subsistence actors are embedded are often dedicated to sharing information (Viswanathan, Sridharan, et al., 2010). In sparsely populated environments, retrieving information from the network is more difficult than it would be in densely populated environments because the number of contacts with other actors is much lower. Actors who adopt a competitor orientation instead can acquire information from “competitors” by observing them at marketplaces, talking to them, and talking about them (e.g., Foster & Rosenzweig, 1995). A well-developed information-sensing competence in turn enables actors to maximize the information extracted from rare moments when they encounter “competitors” and thereby obtain information about customer wants and needs through second- or third-tier network links. This ability may also help to prevent market voids, in which actors lack access to information, and structural holes, which cause actors to rely completely on other actors for information (Burt, 1992). Therefore, in environments with a low population density, a competitor orientation should be beneficial.

**Hypothesis 2.** When the population density is low, a higher competitor orientation leads to a better performance in subsistence markets.

Finally, the positive effect of interfunctional coordination may be challenged by regulative institutional gaps, resource scarcity, and embeddedness. In the absence of formal organizations with departmental boundaries, established procedures, and communication processes, interfunctional coordination must involve two or more network actors that undertake specialized activities to jointly create value for a customer (Viswanathan, Rosa, et al., 2010). These actors may be closely connected through reciprocal processes, as described by social capital theory (Coleman, 1988). Viswanathan, Rosa, et al. (2010) reveal that micro-entrepreneurs distribute resources among customers and vendors while using their family as a buffer to manage these relationships. Tiffern (1995) describes a case study in Kenya in which people meet in shops, bars, and churches to exchange market information, coordinate their activities, and determine their roles to ensure their specialized contribution. This process of interaction, coordination, and specialization requires a minimum level of population density; it cannot function in weakly populated environments. Thus, we hypothesize a positive effect of interfunctional coordination in densely populated areas.

**Hypothesis 3.** When the population density is high, greater interfunctional coordination leads to a better performance in subsistence markets.

5. Methods

5.1. Study context and selection

To test the subsistence market–specific hypotheses on the market orientation–performance relationship, we collected data in pastoralist subsistence markets in Ethiopia (e.g., Galaty & Johnson, 1990). Pastoralists herd livestock such as cattle, camels, goats, and sheep. They have a mobile lifestyle that allows them to utilize remote marginal agricultural lands that are not suitable for permanent crop production but that can temporarily support grazing (Koocheki & Gliessman, 2005), thereby rendering this land economically viable. The scarcity of resources demands this mobile lifestyle, as pastoralists continuously search for water and pasture for their livestock. The cultures are typically characterized by embeddedness and relationships based on reciprocity. Formal institutions are generally weak, and although pastoralists may have formal relationships with local and national governments, they largely govern themselves through traditional structures led by clan elders. Customary rules and laws determine how they manage and use their scarce resources. They manage their herds to fulfill in their own needs and may sell livestock at livestock markets, where they rely on the services of clan brokers who connect them to traders. These traders function in the formal sector and have connections with exporters and slaughterhouses. We use this extreme context as a natural laboratory (Burgess & Steenkamp, 2006) to test the generalizability of the market orientation–performance relationship by applying the recommendations for field research to this subsistence population.

5.1.2. Prestudy

We conducted a prestudy that featured desk and qualitative research. Our desk review involved a survey of related literature on pastoralism and marketing practices. The qualitative prestudy included 138 individual interviews, 14 focus group discussions, and 28 field observations across different regions with (ecologically, demographically, and ethnically) heterogeneous groups of pastoralists, including various livestock marketing chain members: pastoralists, brokers, traders, slaughterhouses, and exporters. We also contacted experts on pastoralism for their opinions and accounts of their experience with field research among pastoralists. These steps helped us to determine the appropriate unit of observation and key respondents, select strata for the sampling
frame, and generate emic items for the market orientation and performance scales. These steps also gave us more confidence in the absence of construct bias by verifying that the market orientation–performance relationship could exist in this context.

5.1.3. Questionnaire development and pretesting
Market orientation research generally focuses on strategic business units as the unit of observation because strategic marketing decisions take place at this level (Webster, 1992). Because pastoralists are not organized into corporate structures, previous studies have used a variety of units of analysis, such as common grazing catchment, village (e.g., Kamara, Swallow, & Kirk, 2003), or household (e.g., Roth, 1991) levels, although individuals do not necessarily make marketing decisions at these levels. Our qualitative pretest revealed that pastoralists make their livestock market decisions at the household level and that the male head of the household is primarily involved in the actual livestock buying and selling. Therefore, we used the household level as our unit of observation and the household head as the key respondent.

We designed a questionnaire in English, which we discussed with an expert who had field research experience in the two focal geographic areas. After incorporating his suggestions, we asked a certified translator to translate the English questionnaire into Oromiffa, the local language spoken in the selected areas. Another translator then back-translated the questionnaire into English to verify the correct interpretation of the questions. Both translators had a thorough understanding of the cultures in which the data were collected.

Surveying pastoralists is challenging because no formal sampling frame or contact information exists, and their mobility makes them difficult to trace (Scarpa, Kristjanson, Drucker, Radeny, Ruto, & Rege, 2001). Therefore, we relied on the services of five government-employed professional interviewers in each area (10 total) to pretest and collect the data. All interviewers were professionally involved with the pastoralists; they maintained relationships between the pastoralists and local governments. They lived in the area of the data collection, shared the cultural background and language of the respondents, and had completed tertiary education. These interviewers received five days of training for data collection. After the training, they conducted two rounds of questionnaire pretesting in which they interviewed 12 pastoralists (two rounds of three respondents in each area).

5.1.4. Sample and interview procedure
In Ethiopia, we selected regions that had received attention from government workers and NGOs and that were sufficiently safe for conducting field research. Yabello in the Southern Borana area and Kereyu Fentale (hereafter, Fentale) in the Rift Valley fit these requirements. These regions also exhibited considerable differences in their population density. The Fentale area had an estimated population density of 70 people per km², whereas Yabello had only 19 people per km² (CSA, 2007). The actual difference is most likely larger, as Fentale is located near the main road between Ethiopia’s capital, Addis Ababa, and its main (foreign) seaport, Djibouti, and because its economic and social life center around this main thoroughway.

Because the two selected areas are very large (Fentale is 1169 km², and Yabello is 5523 km²), we selected specific strata to prevent systematic biases and sampled the pastoralists on a convenience basis. Sampling was conducted as randomly as possible within these strata. Our selection of variables to define strata was informed by the prestudy results. Stratification variables included distance to the main road, participation in additional activities for income (e.g., farming), and conflicts with formal sector activities (e.g., plantations established in areas that had traditionally been used for grazing). With confirmation from local interviewers, we selected four research sites in each area. We found no information suggesting that the strata differed considerably in population size; thus, the number of respondents sampled from each stratum was approximately equal.

Because most pastoralists have low literacy skills, the data were collected through personal interviews that took place in grazing fields. Prior to these visits, we sent messages to the community by contacting community members in the market. The interviewers also explained the purpose of the survey before beginning each interview. Respondents were assured of their anonymity and of the confidentiality of their information. We also explained that the information would be used only for research purposes. However, three and eight respondents from Yabello and Fentale, respectively, refused to participate; the higher refusal number in Fentale reflected the greater suspicion displayed by potential respondents that the results would be used against their interests in disputes with authorities on the right to use certain grazing areas, despite clear indications that the data were being collected for scientific reasons and that the answers were confidential. All interviews in Fentale were conducted prior to 1:00 pm, after which time pastoralists often congregate to chew khat, a stimulant. Interrupting this social event would have been inappropriate. The interviews took an average of one hour; respondents often wanted to explain their answers, and interrupting is not polite in this culture. The explanations also helped interviewers to monitor possible response bias.

We collected data from 130 respondents in Fentale and 102 in Yabello; respondents were approximately evenly distributed across the strata in each area. The number of respondents in Fentale was higher because in the Yabello region, the interviewers were called upon to perform other duties (for example, preparations for national elections and disinfestation of caterpillars). To test for potential long-distance bias, we compared the strata with the longest distance from the road against the other strata for the variables included in the study and several key demographic variables. Only interfunctional coordination showed a significant difference ($M = 1.85$ for the strata near $n = 176$ and $M = 1.51$ for the strata far from the main road $n = 56$; $t_{df=230} = 2.86$, $p < .05$). However, this difference is logical because interfunctional coordination largely relies upon higher population density, as it is found closer to the main roads.

In Fentale, we conducted a second data collection approximately eight months later for two reasons. First, the mobility in this area was seasonally determined, with much less mobility during the rainy season, when the pastoralists have relatively easy access to pasture land and water. The greater mobility in the dry season decreased the population density of the area. For this reason, after conducting the first round of data collection in the dry period, we performed a second round immediately following the rainy season. Second, collecting data for the dependent variable at a later time decreased the potential for single-source bias (Rindfleisch et al., 2008). Using multiple respondents at time 1 was not an option.
because in most households, the oldest son (i.e., the most logical second respondent) left to search for pasture with part of the household’s herd and because interviewing wives was considered inappropriate. Even during the rainy season, it was challenging to find pastoralists within a reasonable distance. Therefore, we managed to obtain data at t2 from 71 of the original 130 respondents (55%). We compared the respondents at t2 with non-respondents at t2 by using two-group t-tests on the variables included in the study and demographics. None of these variables exhibited significant differences; thus, a selection bias due to mobility appeared unlikely.

5.2. Operationalization and measurement

By using the transcripts from the qualitative prestudy, the first and second authors each identified a list of items to measure customer orientation, competitor orientation, and interfunctional coordination. We compared these lists and discussed any differences; we used the same technique to identify performance measures and the control variable (perceived rainfall). To graphically depict the response categories, we replaced a traditional five-point Likert-type scale with five sticks of increasing worth, such that stick 5 was worth five times as much as stick 1. As shown in Fig. 2, pastoralists responded to the multi-item questions by selecting sticks (stick 1 = “strongly disagree,” stick 5 = “strongly agree”). To ensure that respondents understood the procedure, they practiced prior to the start of the interview.

To validate the multi-item measures, we used conventional methods, including exploratory factor analysis and Cronbach’s alpha (Churchill, 1979). We dropped any items that loaded on multiple factors or that had low loadings. Because we made no direct comparison between groups, we included the three market orientation constructs in a single CFA model by using LISREL 8.7 (Jöreskog & Sörbom, 2004).1

After the purification step, we measured customer orientation by using five items that refer to practices valued by buyers, such as improved breeding (alpha = .86). These items jointly reflect business practices that require an understanding of customers and an aim to satisfy their needs by creating customer value (Narver & Slater, 1990). The measure of competitor orientation included four items pertaining to the collection of and response to information about other suppliers in the livestock market (alpha = .77). For interfunctional coordination, we measured the amount of rain, as perceived by a pastoralist, with six items (alpha = .78). Pastoralists depend heavily on rain to support their livestock (Smith, Barrett, & Box, 2001); thus, we controlled for performance differences due to rain. Finally, we measured population density with a dummy variable that distinguished between the two areas (1 = Yabello, 0 = Fentale). Table 3 contains the means, standard deviations, and correlations of the variables.

5.3. Data analysis

We tested our hypotheses by using ordinary least squares regression models. Specifically, we applied the following model (Model 1) to analyze the data collected at t1:

\[
Y = \beta_0 + \beta_1X1 + \beta_2X2 + \beta_3X3 + \beta_4X4 + \beta_5X5 + \beta_6X6 + \beta_7X7 + \beta_8X8 + \beta_9X9 + \epsilon, \]

where Y is performance (dependent variable); \(\beta_s\) are the parameter estimates; and X1–X3 denote customer orientation, competitor orientation, and interfunctional coordination, respectively. In turn, X4 corresponds to the area dummy; X5–X7 correspond to the three control variables: family size, market interaction, and rainfall, respectively; and X8 and X9 correspond to the multiplicative interaction terms between the area dummy and either competitor orientation or interfunctional coordination. Models 2 and 3 provide the separate estimates for Yabello and Fentale, respectively, which we used to uncover any within-country heterogeneity (Burgess & Steenkamp, 2006). Model 4 also reports the results for Fentale; however, it uses the performance measure at t2, shortly after the rainy season. Except for the dummy, we mean-centered the independent variables before inputting them (Aiken & West, 1991). We also inspected our findings for multicollinearity; however, because the highest variance inflation factor in our models was 3.49 (competitor orientation, Model 1), multicollinearity did not appear to be a problem for our analyses (Hair, Anderson, Tatham, & Black, 1995).

6. Results

We present the study results in Table 4. We predicted a positive effect of customer orientation, and the effect of customer orientation on performance is significant in Models 1–3 at t1 and in Model 4 at t2. Thus, we found strong support for Hypothesis 1; the first three models show that the effect is consistent across both areas, and
comes stronger, just as our theory predicted. The most likely explanation for this finding is that creating a multiplicative term between the area dummy for population density and competitor orientation in Model 1 is not significant, nor is the effect of competitor orientation significant in Model 2. We could not test this hypothesis at \( t2 \), which included data only from the more densely populated Fentale area. However, in line with the underlying rationale for this hypothesis, we note that the effects of competitor orientation in Fentale are not significant at \( t1 \) or \( t2 \) (Models 3 and 4). Hypothesis 3 predicted a positive effect of interfunctional coordination on performance with a high population density; we tested this hypothesis by examining the effect of interfunctional coordination in the densely populated Yabello area. In Model 1, the multiplicative term between interfunctional coordination and the area dummy is positive and significant (\( p<.1 \)), in support of Hypothesis 3. The same effect appears in Model 3, showing a positive and significant main effect of interfunctional coordination on performance. In contrast, this main effect is not significant in the sparsely populated Yabello area (interfunctional coordination \( \beta = .001 \) in Model 2). This finding presents strong support for Hypothesis 3 because we obtained this evidence during the dry season, when the population density in the Fentale area was even lower due to the pastoralists' greater mobility. In Model 4, the data gathered shortly after the rainy season indicate a higher effect size of interfunctional coordination (Model 4, \( \beta = .054 \), \( p<.01 \); Model 3, \( \beta = .149 \), \( p<.05 \)). When population density is higher, the effect of interfunctional coordination on performance becomes stronger, just as our theory predicted.

### Table 3

Means, standard deviations, and correlations.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>SD</th>
<th>Yabello Mean</th>
<th>SD</th>
<th>Fentale Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Customer orient.</td>
<td>3.01</td>
<td>1.09</td>
<td>2.67</td>
<td>1.30</td>
<td>3.35</td>
<td>1.72</td>
</tr>
<tr>
<td>2 Competitor orient.</td>
<td>3.67</td>
<td>.98</td>
<td>3.02</td>
<td>.90</td>
<td>4.18</td>
<td>.69</td>
</tr>
<tr>
<td>3 Interfunc. coord.</td>
<td>1.77</td>
<td>.78</td>
<td>1.58</td>
<td>.85</td>
<td>1.31</td>
<td>.89</td>
</tr>
<tr>
<td>4 Perceived rainfall</td>
<td>1.67</td>
<td>.62</td>
<td>1.70</td>
<td>.67</td>
<td>1.64</td>
<td>.57</td>
</tr>
<tr>
<td>5 Family size</td>
<td>8.10</td>
<td>4.34</td>
<td>8.61</td>
<td>4.12</td>
<td>7.70</td>
<td>4.48</td>
</tr>
<tr>
<td>6 Mkt. interaction</td>
<td>21.59</td>
<td>15.58</td>
<td>18.47</td>
<td>13.49</td>
<td>23.88</td>
<td>16.81</td>
</tr>
<tr>
<td>7 Area dummy</td>
<td>.56</td>
<td>.50</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>8 Performanc1</td>
<td>3.14</td>
<td>.87</td>
<td>2.95</td>
<td>.89</td>
<td>3.29</td>
<td>.82</td>
</tr>
<tr>
<td>9 Performance t2</td>
<td>3.39</td>
<td>.83</td>
<td>N.A.</td>
<td>N.A.</td>
<td>3.39</td>
<td>.83</td>
</tr>
</tbody>
</table>

N.A. = not applicable.

\( a \ p<.01 \)

\( b \ p<.05 \)

\( c \ p<.01 \), two tailed significance.

### Table 4

Standardized regression coefficients for market orientation components on performance.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t1, \text{both regions (N=232)} )</td>
<td>( t1, \text{Yabello (N=102)} )</td>
<td>( t1, \text{Fentale, (N=130)} )</td>
<td>( t2, \text{Fentale (N=71)} )</td>
</tr>
<tr>
<td>Beta</td>
<td>( t )-Value</td>
<td>Beta</td>
<td>( t )-Value</td>
</tr>
<tr>
<td>Customer or.</td>
<td>.362</td>
<td>5.08</td>
<td>.303</td>
</tr>
<tr>
<td>Competitor or.</td>
<td>-.039</td>
<td>-.35</td>
<td>-.020</td>
</tr>
<tr>
<td>Interf. coordination</td>
<td>-.022</td>
<td>-.25</td>
<td>.001</td>
</tr>
<tr>
<td>Area dummy</td>
<td>.086</td>
<td>1.07</td>
<td>.147</td>
</tr>
<tr>
<td>Family size</td>
<td>.115</td>
<td>1.81</td>
<td>.12</td>
</tr>
<tr>
<td>Market interaction</td>
<td>-.007</td>
<td>-.12</td>
<td>.012</td>
</tr>
<tr>
<td>Perceived rainfall</td>
<td>.121</td>
<td>1.85</td>
<td>.219</td>
</tr>
<tr>
<td>F-statistics (df)</td>
<td>F(9, 222), 5.25</td>
<td>F(6, 95), 3.19</td>
<td>F(6, 123), 3.97</td>
</tr>
<tr>
<td>( R^2 ) (adj., ( R^2 ))</td>
<td>.176 (.142)</td>
<td>.168 (.115)</td>
<td>.162 (.122)</td>
</tr>
</tbody>
</table>

N.A. = not applicable.

\( a \ p<.01 \)

\( b \ p<.05 \)

\( c \ p<.10 \), one-tailed significance.

### 7. Discussion and Implications

Subsistence markets provide a context that challenges marketing theories and research practices. Starting from a discussion on field research in subsistence populations, we have compiled a list of challenges that impact bias and equivalence in subsistence markets and have formulated recommendations for effective marketing research practices in these contexts. To illustrate the challenge of testing marketing theories in subsistence settings, we conducted a study of the market orientation–performance relationship among the pastoralist population of Ethiopia. We found that customer orientation—the core feature of a market orientation–led to superior performance across both areas studied and in two time periods. The finding implies that creating value for customers should be a primary concern in subsistence markets as much as in high-income markets. However, the positive effect of interfunctional coordination is contingent on network density as a specific feature of subsistence markets. Whereas in high-income markets, such density can be created by bringing different business functions together in bounded organizations, subsistence markets depend on the network context in which micro-entrepreneurs relate to one another. This finding indicates that previous results with respect to network effects in urban subsistence markets (e.g., Viswanathan, Sridharan et al., 2010) may not be fully generalizable to low-density contexts in rural areas.

Finally, our data do not support a positive effect of competitor orientation on performance. The most likely explanation for this finding stems from our qualitative prestudy, in which several informants mentioned their use of mobile phones to obtain market information. For these respondents, it had become less important to monitor how other pastoralists approached the market. Still, this finding is not exceptional; other studies have questioned the added value of a
Table 5
Recommendations for field research in subsistence populations.

Prestudy
- Insights on the subsistence population should be obtained through expert interviews, field interviews and observations, and desk and literature research. Insights should be obtained on:
  - The meaning of and relationships between marketing concepts.
  - The possibility of stratified sampling and the relevant strata (and relative sizes) in the research population.
  - The unit of analysis and key respondents.
  - Potential items and measures used in other academic disciplines.
  - Potential social or cultural barriers to get access to respondents and ways to solve them (e.g., involving nonprofit organizations).
  - Potential physical barriers to reach respondents (and consequences for the generalizability of leaving out areas that are difficult to reach).
  - If longitudinal studies are conducted, seasonal influences and potential dropouts due to mobility.
  - Potential problems in obtaining information due to cultural, class, or gender differences.
  - Whether and how respondents can be isolated from group members, if individual or multirespondent data are desirable.
- If secondary information is used for sampling decisions, the information should be carefully inspected in terms of reliability and validity.

Questionnaire development and pretesting
- If respondents are low-literate, design the questionnaire so that it suits personal interviews in a language that respondents are comfortable with.
- Adjust existing measures and stimuli so that they fit the local context and are suitable for concrete thinking, make use of pictographic thinking, and avoid complex numerical tasks.
- Avoid questions that are perceived as sensitive in the local culture, or ask these questions at the end of the interview, after trust has been developed between the respondents and the interviewer.
- If past decisions are examined, assess the importance of the decision and the time since the decision was made to investigate potential recall biases.
- If stratified sampling is used, include selection questions for studies conducted in urban or other “mixed” environments to determine in which stratum respondents fit.
- Pretest questionnaires in terms of interpretation and sensitivity with experts and in all targeted strata or clusters.
- Involve research assistants and/or interpreters in the pretest as part of their training.

Data collection and analysis
- If transportation systems are weak, put more time and effort into data collection, or test for a potential response bias between easy- and difficult-to-reach respondents.
- Select interviewers (and additional interpreters if needed) on the basis of their level of education and understanding of local cultures and languages.
- Train interviewers in a systematic way (especially if multiple interviewers are used), including recognition of response patterns; supervise and observe interviewers during interviews; inspect completed questionnaires; and give immediate feedback about possible errors.
- Provide information to respondents about the study objective, affiliations of the researcher and interviewer, and anonymity of their responses well before the interview takes place. Thus respondents can discuss their possible participation with others or ask for approval.
- Report the number of potential respondents that refuse to participate.
- Rewards should be dependent on the social capital context and used with caution, especially for snowball sampling.
- Clearly explain to respondents that there are no right or wrong answers in the interview.
- Check (and if possible correct) for possible biases, especially when not all a priori measures to prevent the bias have been applied.

Sevemseveral limitations of the empirical illustration also deserve attention. Although we systematically selected the study context, these data reflect the pastoralist context in one country; therefore, the findings may be context and country specific. Our study relies mainly on cross-sectional data, although we also collected performance data in one of the regions at a second time period. Panel data may reveal further consequences of market orientation.

In sum, the validity and reliability of our measures, the support for our hypotheses, and the reasonable explanations for the unsupported hypothesis together provide confidence in the methods we used to address the specific field research challenges of subsistence settings. We have demonstrated that by taking the specific context of subsistence markets into consideration, researchers can formally test the generalizability of marketing theories even in environments that are institutional-ly remote from the high-income markets in which the theories were developed. Addressing bias and equivalence is important in any generalizability study; however, the institutional distance between subsistence and high-income markets renders researchers from high-income markets particularly susceptible to overlooking the sources of bias that are typical of subsistence markets. Preventing bias by following appropriate methodological practices therefore deserves close attention in the design and data collection stages of the research process. These practices are summarized in Table 5. A posteriori procedures may subsequently test the assumptions regarding bias and equivalence after data collection. Marketing researchers in subsistence markets can build upon recent methodological contributions, such as IRT, that help to develop contextual, yet comparable, measures; identify response styles; and reduce item bias. In these and other respects, advanced a posteriori methodologies can further mine the insights obtained from subsistence markets and can help to test the generalizability of marketing theories.

In summary, we encourage further quantitative research regarding subsistence markets. Such research should take careful account of the specific challenges inherent to this type of research with respect to bias and equivalence. Moreover, it should consider our recommendations for addressing these challenges. Such quantitative research has great potential to generate practically, theoretically, and socially relevant findings regarding the generalizability of mainstream marketing theory.

Acknowledgments

Financial support for this study was obtained through a grant to the second author by the joint project NPT-ETH-152 between the Netherlands Organization for International Cooperation in Higher Education and Research (Nuffic) and Haramaya University of Ethiopia. The authors wish to thank the interviewers who assisted with the field work in Ethiopia. The authors are grateful to the special issue editors and two anonymous reviewers for their helpful comments on previous drafts of this article.

Appendix A. Construct items, loadings, and alpha values

<table>
<thead>
<tr>
<th>Item</th>
<th>Customer orientation (alpha = .86, eigenvalue = 3.18)</th>
<th>Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>We do nothing to increase the quality of our livestock that we want to sell. (R)</td>
<td>.76</td>
</tr>
<tr>
<td>2</td>
<td>We increase the quality of our livestock that we are planning to sell in the market.</td>
<td>.88</td>
</tr>
<tr>
<td>3</td>
<td>We breed with livestock that will give us the quality traders are looking for.</td>
<td>.83</td>
</tr>
<tr>
<td>4</td>
<td>We always prefer to keep the best livestock for ourselves. (R)</td>
<td>Dropped</td>
</tr>
<tr>
<td>5</td>
<td>We sell our livestock only when we could not get income from other sources. (R)</td>
<td>Dropped</td>
</tr>
<tr>
<td>6</td>
<td>We raise livestock that the market wants.</td>
<td>.70</td>
</tr>
<tr>
<td>7</td>
<td>We always search for better breeds to satisfy traders and exporters.</td>
<td>.81</td>
</tr>
</tbody>
</table>
Arnould, E. J. (1989). Toward a broadened theory of preference formation and the
Anker, R. (1983). Female labour force participation in developing countries: A critique
References
Appendix A (continued)

<table>
<thead>
<tr>
<th>Item</th>
<th>Interfunctional coordination (alpha = .79, eigenvalue = 2.85) Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What other livestock suppliers are doing in the market does not bother to me. (R)</td>
</tr>
<tr>
<td>2</td>
<td>We always check what other livestock suppliers are doing on the mark</td>
</tr>
<tr>
<td>3</td>
<td>Knowing the livestock type that others are supplying to the market is important to us.</td>
</tr>
<tr>
<td>4</td>
<td>We always decrease or increase our market price following other suppliers.</td>
</tr>
<tr>
<td>5</td>
<td>We are not interested in what other pastoralists are doing in the market.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Interfunctional coordination (alpha = .79, eigenvalue = 2.85) Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Our broker will tell us when prices for our livestock are good.</td>
</tr>
<tr>
<td>2</td>
<td>We collaborate very closely with our broker.</td>
</tr>
<tr>
<td>3</td>
<td>Our broker advises us for best breed and fattening to increase quality of our livestock.</td>
</tr>
<tr>
<td>4</td>
<td>Brokers withhold important market information from us. (R)</td>
</tr>
<tr>
<td>5</td>
<td>We talk to community members on how to improve the quality of our livestock.</td>
</tr>
<tr>
<td>6</td>
<td>We exchange information in the community before going to the market.</td>
</tr>
<tr>
<td>7</td>
<td>We always contact knowledgeable people (e.g., experts) for market information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Interfunctional coordination (alpha = .79, eigenvalue = 2.85) Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>We can rely on the rain.</td>
</tr>
<tr>
<td>2</td>
<td>Rain comes always as we expected.</td>
</tr>
<tr>
<td>3</td>
<td>These days rain is not coming as we expected. (R)</td>
</tr>
<tr>
<td>4</td>
<td>Unexpected droughts may happen to us. (R)</td>
</tr>
<tr>
<td>5</td>
<td>Rain always comes at the same time of the year.</td>
</tr>
<tr>
<td>6</td>
<td>We assumed that rain would come but it didn’t. (R)</td>
</tr>
<tr>
<td>7</td>
<td>We are getting less rainfall than we expected. (R)</td>
</tr>
</tbody>
</table>

Performance

<table>
<thead>
<tr>
<th>Item</th>
<th>Time 1 (alpha = .70, eigenvalue = 2.14) Factor loadings</th>
<th>Time 2 (alpha = .63, eigenvalue = 1.97) Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Increasing the herd size.</td>
<td>Dropped</td>
</tr>
<tr>
<td>2</td>
<td>Education of my children.</td>
<td>.81</td>
</tr>
<tr>
<td>3</td>
<td>Extra income generated from livestock production.</td>
<td>.75</td>
</tr>
<tr>
<td>4</td>
<td>Saving money.</td>
<td>Dropped</td>
</tr>
<tr>
<td>5</td>
<td>Diversifying different activities (e.g., petty trading),</td>
<td>.52</td>
</tr>
<tr>
<td>6</td>
<td>Growing crops in addition to livestock raising.</td>
<td>.74</td>
</tr>
</tbody>
</table>

Notes: (R) indicates a reversed item.


Competitive information, trust, brand consideration and sales: Two field experiments

Guilherme Liberali a, b, *, Glen L. Urban b,1, John R. Hauser c,1

a Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR, Rotterdam, The Netherlands
b MIT Sloan School of Management, Massachusetts Institute of Technology, E52-536, 77 Massachusetts Avenue, Cambridge, MA 02139, United States
c MIT Sloan School of Management, Massachusetts Institute of Technology, E52-538, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

A R T I C L E   I N F O

Article history:
First received in 3, April 2012 and was under review for 2 months
Available online 17 October 2012
Area Editor: Gary L. Lilien

Keywords:
Competitive information
Brand consideration
Electronic marketing
Information search
Signaling
Trust

A B S T R A C T

Two field experiments examine whether providing information to consumers regarding competitive products builds trust. Established theory suggests that (1) competitive information leads to trust because it demonstrates the firm is altruistic, and (2) trust leads to brand consideration and sales. In year 1, an American automaker provided experiential, product-feature, word-of-mouth, and advisor information to consumers in a 2 × 2 × 2 × 2 random-assignment field experiment that lasted six months. Main-effect analyses, conditional-logit models, and continuous-time Markov models suggest that competitive information enhances brand consideration and possibly sales and that the effects are mediated through trust. However, in a modification to extant theory, effects are significant only for positively valenced information. The year-2 experiment tested whether a signal that the firm was willing to share competitive information would engender trust, brand consideration, and sales. Contrary to many theories, the signal did not achieve these predicted outcomes because, in the year-2 experiment, consumers who already trusted the automaker were more likely to opt in to competitive information. In addition to interpreting the field experiments in light of extant theory, we examine cost effectiveness and describe the automaker’s successful implementation of revised competitive-information strategies.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction and motivation

In 2003–2005, an American automaker (“AAM”) faced a situation common among firms that had once dominated their industries. New competitors had entered with products perceived to be higher in quality and better matched to consumer needs. The automaker’s brands no longer connoted quality or innovation; brand strength had declined. By 2003, over half of the consumers in the U.S. (and almost two-thirds in California) would not even consider AAM brands; its brands were effectively competing for only a fraction of the market. Ingrassia (210, p. 163) cites the lack of brand consideration as a cause of the decline of American brands. Although the automaker had recently invested heavily in design and engineering, the automaker would never again regain strength nor market share unless its brands could earn consumers’ brand consideration.

The situation of AAM was similar to that faced by many once-dominant brands. After the financial crises of 2008, consumers were hesitant to consider and purchase financial services from established firms such as New York Life (Green, 2010). Research in Motion (Blackberry), Motorola, and Nokia once dominated the smart phone market, but lost market share and brand image to Apple’s iPhone. Many consumers reject these once-dominant firms as no longer relevant. Firms entering dominated markets also face the challenge of earning consumers’ brand consideration. Examples include Barnes & Noble’s Nook (after Amazon’s Kindle) and the Kindle Fire (after Apple’s iPad). In other situations, a firm might stumble, perhaps through no fault of its own, and face an uphill battle to be considered as a viable alternative again. Examples include Tylenol after the 1982 poisonings, the Audi Quattro after the 1986 publicity on sudden acceleration, Toyota after widespread media reports in 2010 that software problems led to sudden acceleration, Toyota after widespread media reports in 2010 that software problems led to sudden acceleration, and Genentech after production problems in critical drugs. In B2B markets, the Brazilian aircraft manufacturer Embraer launched a new line of well-designed and high-quality business jets, but found it hard to enter buyers’ consideration sets when competing with such established firms as Bombardier, Dassault, Cessna, Hawker Beechcraft, and Gulfstream (Aboulafia, 2007).

Many authors recommend that firms earn brand consideration by first earning trust from consumers. If consumers trust a brand, then
consumers are motivated to invest the time and effort to learn more about the features of the brands (see, e.g., Urban, 2005 and citations therein). Research suggests that to earn trust a firm should make itself vulnerable by acting altruistically and putting consumers’ needs above its own (Kirmani & Rao, 2000; Moorman, Deshpandé, & Zaltman, 1993; Rousseau,Sitkin, Burt, & Camerer, 1998). For example, New York Life stated publicly that “the guarantees we make last a lifetime” and that “our promises have no expiration date” (Green, 2010). Embraer, primarily a manufacturer of commercial regional jets, took a risky strategy of announcing a new line of jets specifically designed for business travel (Aboulafia, 2007). Even Procter & Gamble’s Global Marketing Officer began new initiatives because “market share is trust materialized” (Bloom, 2007).

The theory of making oneself vulnerable is persuasive, but for a major automaker, such a strategy puts billions of dollars at risk. Most evidence to date is based on cross-sectional surveys (structural equation models and meta-analyses) or laboratory experiments. For example, in a review of the literature, Geyskens, Steenkamp, and Kumar (1998) cite 20 studies based on surveys, seven based on laboratory experiments, and none based on field experiments. Even today, we find few field experiments and none in which a once-dominant firm made itself vulnerable to gain trust. Before committing to a trust-based strategy, AAM decided that the strategy had to be proven in rigorous field experiments. We were interested in partnering because we felt the need to field-test theories that were developed from surveys and laboratory experiments.

Specifically, AAM would signal that it was acting in the consumers’ best interests by providing unbiased competitive information to its customers. If the theories were correct, consumers would come to trust AAM and consider its brands. The vulnerability signal was risky because AAM would not win all head-to-head comparisons with its competitors and because goodwill was lacking among consumers familiar with pre-2003 products. However, if the theories were correct, competitive information would signal trust and have a positive impact. AAM’s post-2003 products would win on sufficiently many comparisons that brand consideration would overcome losses due to adverse short-term comparisons.

In this paper, we describe the field experiments and general lessons. The year-1 data suggest that competitive information leads to trust, brand consideration, and sales but that the effect of competitive information is mediated through trust. However, contrary to extant theory, these effects are significant only for positively valenced information. The year-2 data suggest that contrary to many theories, a signal alone, indicating that the firm is willing to share competitive information, does not engender trust, brand consideration, and sales. The signal was not effective because consumers who already trusted the automaker were more likely to opt in to competitive information. Managerial analyses suggest that competitive information is effective when targeted cost-effectively to consumers who are otherwise skeptical of the brand. We begin with the theory.

2. Underlying theory

There is considerable precedent in the trust and signaling literatures to support the theory that providing competitive information would enhance trust, leading to brand consideration and sales.

2.1. Competitive information, vulnerability, and trust

Morgan and Hunt (1994) provide evidence that commitment leads to trust and that the sharing of meaningful and timely information is an antecedent of trust. This classic article was of particular interest because the Morgan–Hunt data were based on independent automobile tire retailers. However, the theories cut across industries. In a study of market research suppliers, Moorman et al. (1993) suggest that sincerity, timeliness, and the willingness to reduce uncertainty (presumably by sharing information) all lead to trust. Indeed, vulnerability to gain trust is a common theme throughout the literature (Kirmani & Rao, 2000; Moorman et al., 1993; Rousseau et al., 1998). Other researchers define trust with the related concepts of credible information and acting benevolently toward the other party (Doney & Cannon, 1997; Ganesan, 1994; Ganesan & Hess, 1997; Geyskens et al., 1998; Gurviez & Korchia, 2003; Maister, Green, & Galford, 2000; Sirdeshmukh, Singh, & Sabol, 2002; Urban, 2004).

Providing competitive information to achieve vulnerability, credibility, and an image of altruism is a common recommendation (Trifts & Häubl, 2003; Urban, 2004; Urban, Amyx, & Lorenzon, 2009; Urban, Sultan, & Qualls, 2000). Bart, Shankar, Sultan, and Urban (2005) provide evidence in online settings that clear, unbiased navigation and presentation enhance trust. Shankar, Urban, and Sultan (2002) argue that quality and timeliness of information enhances trust. In laboratory experiments, Trifts and Häubl (2003) show that competitive price information enhances an online retailer’s perceived trustworthiness. These authors go on to state, citing other authors, that “the sharing of relevant and potentially self-damaging information can be viewed as a type of communication openness, which is an important form of trust-building behavior and has been found to have a direct positive relationship with the level of trust” (p. 151).

However, gaining trust is only useful if it leads to sales. Fortunately, there is ample evidence supporting this conclusion. Bart et al. (2005) show that trust mediates relationships between website design characteristics and purchase intentions. Crosby, Evans, and Cowles (1990) show that enhanced trust leads to continued exchanges between parties and to sales. Doney and Cannon (1997) demonstrate that trust influences buyer’s anticipated future interactions. Ganesan (1994) shows that trust and interdependence determine the long-term orientation of both retail buyers and their vendors. Geyskens et al. (1998), in a meta-analysis of 24 papers (27 studies), suggest that trust mediates roughly 48% of long-term outcomes. Trifts and Häubl (2003) show that the effect of competitive price information is mediated through trust. Hoffman, Novak, and Peralta (1999, p. 85) posit that “the most effective way for commercial web providers to develop profitable exchange relationships with online customers is to earn their trust.” Sirdeshmukh et al. (2002) provide evidence that benevolent management policies and procedures enhance trust and that trust enhances long-term loyalty. Büttner and Göritz (2008) provide evidence that trust enhances intentions to buy.

Based on this literature, we posit that sharing competitive information with consumers will demonstrate that AAM is both altruistic and vulnerable. Altruism and vulnerability will cause consumers to trust the automaker, which, in turn, will lead to brand consideration and sales. We also posit that the effect of competitive information will be mediated through trust. The following hypotheses are tested in the year-1 field experiment.

H1. Competitive information provided to the test group will increase consumers’ trust of the automaker relative to the control group.

H2. Competitive information provided to the test group will increase brand consideration relative to the control group.

H3. Conditional on brand consideration, competitive information provided to the test group will increase sales relative to the control group.

H4. The increase in brand consideration and sales in the test group is mediated through trust.

The literature does not distinguish between positively valenced and negatively valenced competitive information: both are assumed to communicate altruism and vulnerability and hence lead to trust, brand consideration, and sales. However, we know that negative information per se decreases brand consideration and sales relative to positive information (e.g., automotive experiments by Urban, Hauser, & Roberts, 1990, p. 407). Thus, we expect that H1 through H3 are more likely to be supported for positively valenced competitive information than for
negatively valenced competitive information. A priori, we do not know the relative strengths of the negative information and the altruism vulnerability effects; thus, it becomes an empirical question whether H1 through H3 are supported for negatively valenced information.

2.2. Trust as a signal

Automotive competitive information comes in many forms, varying from a simple list of competitive features to community forums or online advisors, all the way to providing consumers the ability to test drive competitive vehicles. Some data are inexpensive (simple lists) and some are quite costly (competitive test drives). We would therefore like to disentangle the effect of actually providing competitive information from the act of sending a vulnerability (trust) signal by offering to provide competitive information.

Kirmani and Rao (2000) provide a comprehensive review of the literature on signaling unobserved product quality. These researchers argue that revenue-risking signals work when consumers can infer that it is in the long-term revenue interests of high-quality firms to provide the signal but not in the long-term revenue interests of low-quality firms to do so. The theories require that (1) information is hard to obtain pre-purchase, (2) quality can be assessed post-purchase, and (3) that the “bond” of vulnerability (supplied in equilibrium only by high-quality firms) is credible. Condition 2 is clearly met in automotive markets. Condition 1 is met for expensive and difficult-to-obtain information such as competitive test drives, but it may not be met for simple lists of features. Condition 3 requires that the competitive information is expensive to provide so that consumers do not interpret the information as “cheap talk.” This condition is met for the competitive information provided in AAM’s experiments. Only an automaker with post-evaluation high-quality brands would want to risk providing competitive information that was expensive and otherwise hard to obtain.

Signaling theory is extensive, beginning with Spence’s (1973) classic article. See also Milgrom and Roberts (1986). In marketing, Biswas and Biswas (2004, p. 43) provide experiments that “signals are stronger relievers of risk in the online setting, especially for products with high non-digital attributes.” Erdem and Swait (1998, p. 137) argue that the brand itself can be a signal, especially if “the information about a brand’s position that is communicated to the consumer by a firm [is] perceived as truthful and dependable.” Erdem and Swait (2004) argue explicitly that brand credibility, operationalized as trustworthiness and expertise, is a signal that increases brand consideration.

Given the strong support for a signaling theory of trust and the fact that the willingness to accept vulnerability by providing competitive information could signal that an automaker is a high-quality, trusted brand, we state the following hypotheses that we sought to test in the year-2 field experiment:

H5. A signal that competitive information is readily available increases trust in the test group relative to the control group.

H6. A signal that competitive information is readily available increases brand consideration in the test group relative to the control group.

H7. Conditional on brand consideration, a signal that competitive information is readily available increases sales in the test group relative to the control group.

H8. If H5 is true and if either H6 or H7 is true, then the increase in brand consideration and/or sales due to the signal are mediated through trust (H8 is conditional on whether H5 through H7 are supported).

2.3. Types of competitive information

Urban et al. (2000) suggest that virtual advisors and complete-and-unbiased information on competitive products are important to building trust. Urban et al. (2009) suggest that firms provide honest open advice and information on competitive offerings. Häubl and Trifts (2000) provide evidence that virtual recommendation agents (advisors) and comparative product information assist consumers in brand consideration. Arora et al. (2008) suggest further that information sources be personalized by the firm to the consumer (as in customized brochures) or capable of being customized by the consumer (as in the availability of drive experiences for a wide range of vehicles). We (and AAM) wanted to test different types of competitive information. The generic types were chosen to span the range of information available in automotive markets (for managerial decisions made at AAM) and as representing generic types of information that would be available in other categories (for generality). To avoid cheap talk, the information sources (in year 1) were all expensive for AAM to provide and included the following:

- Direct product experience. In automotive markets, this means test drives or their equivalent, such as renting a car or truck. Unlike other information sources, no automaker (at the time) provided competitive test drives (auto shows were a poor substitute). Direct product experience is typical in other high involvement (but less expensive) categories. For example, Proctor & Gamble routinely sends out product samples and encourages consumers to do direct comparisons with currently used brands, and exercise equipment manufacturers place their products in fitness centers to allow consumers to try them.
- Print and online information. In automotive markets, this takes the form of competitive brochures or information abstracted from those brochures. In other markets, such information is available by searching product catalogs or from information aggregation websites, such as CNET in electronic goods.
- Word-of-mouth. In automotive markets (2003–2005), word-of-mouth information on competitors was available through online automotive communities. This type of information is a surrogate for information now available in Angie’s List, Cyworld, Facebook, Google+, Yelp, and other social networks.
- Trusted advisors. Many websites provide unbiased online advisors (and some biased advisors). Such advisors are available in many product categories.

3. Year 1: randomized experiments for competitive information

The year-1 field experiment tested Hypotheses H1 through H4, whether competitive information leads to trust and trust leads to brand consideration and sales. To ensure that the year-1 field experiment tests competitive information and not just the signal that competitive information is available, consumers in year 1 are given incentives to experience the competitive information. The year-2 field experiment tested Hypotheses H5 through H8, whether the sheer act of offering competitive information signals vulnerability and altruism which in turn leads to trust, brand consideration, and sales. In year 2, competitive information was made available, but consumers experienced the information only if they opted in.

3.1. Consumer panel observed over six months

The year-1 panel ran monthly from October 2003 to April 2004 (this was five years prior to the bankruptcies of two American automakers, both of which are now profitable). Members of Harris Interactive’s panel were screened to be in the market for a new vehicle in the next year, on average within the next 6.6 months, and were invited to participate and complete six monthly questionnaires. In total, Harris Interactive enrolled 615 Los Angeles consumers of whom 317 completed all six questionnaires for an average completion/retention rate of 51.5%. We were unable to obtain exact recruitment rate statistics for year 1, but Harris Interactive estimates an initial recruitment rate of approximately 40%.

G. Liberali et al. / Intern. J. of Research in Marketing 30 (2013) 101–113
Consumers were assigned randomly to four experimental treatments, each representing one of the four generic forms of competitive information. In year 1 (but not year 2), consumers assigned to a treatment were given sufficient incentives to experience that treatment. Treatment were assigned in a fully crossed orthogonal design giving rise to a 2 × 2 × 2 × 2 full-factorial field experiment such that various respondents received 0, 1, 2, 3, or 4 treatments. Although assignments were random with a goal of 50–50 assignment, the logistics were such that only approximately 40% of the panel members were randomly assigned to competitive test drives. The other treatments were close to 50–50. By design, the competitive online advisor was available in all months, the competitive community ran for all but the last month, the customized brochures were mailed in months 2 and 3, and the competitive test drive took place in month 4 (this differential assignment is analyzed with models that take account of the differential timing of assignments). The exact numbers of consumers assigned to each treatment in year 1 is summarized in Table 1.

### 3.2. Competitive information experimental treatments

All experimental treatments were produced and managed professionally and required substantial investments ranging from approximately $150,000 (brochures) to $1 million (competitive test drives). To avoid cheap talk, all treatments would be quite expensive on a national basis. Direct product experience was represented by a test drive experience at a California test track in which consumers could drive vehicles from Acura, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ford, Honda, Lexus, Lincoln, Mercedes, Pontiac, Saab, Saturn, Toyota, Volkswagen, and Volvo without any sales pressure (Fig. 1a).

Print and online information was represented by customized brochures (year 1) and competitive brochures (year 2). Customized brochures were glossy brochures tailored to the needs of individual consumers and mailed directly to the consumers (Fig. 1b). Competitive brochures were less customized, web-based, and included brochures from all competitors. While the year-1 print information was not competitive information, AAM believed it would engender trust and signal altruism. All other information is competitive.

Word-of-mouth information was represented by an unbiased online CommuniSpace™ forum in which consumers could participate in over 30 discussions about both AAM and competitive vehicles (Fig. 1c). Commensurate with concerns that such information might make the automaker vulnerable, approximately 20% of the comments about AAM brands were negative.

Trusted advisors were represented by an online advisor that was co-branded with Kelley Blue Book and similar to that developed by Urban and Hauser (2004) (see Fig. 1d). In year 1, the online advisor recommended competitors approximately 83% of the time.

### 3.3. Dependent measures: brand consideration and purchase

Brand consideration was measured with drop-down menus in which the consumer indicated the brands that he or she would consider. Brand consideration is a quantal measure (consider vs. not consider) where a consumer is coded as considering AAM if the consumer indicated that he or she would consider one of AAM’s brands. To avoid demand artifacts, AAM was not identified as the sponsor of the study and AAM vehicles occurred throughout the list of potential vehicles. Because consumers might evaluate and reject a brand, they can report no brand consideration in month t, even if they considered a brand in month t − 1.

Purchase (sales) is also a quantal measure (purchase or not purchase) measured with a standard stated purchase measure. Once consumers purchase a vehicle, we assume they cannot un-consider and un-purchase that vehicle during the six-month observation period.

### 3.4. Trust

We hypothesize that trust is a dependent measure and a mediator for brand consideration and purchase (trust is a firm-specific property in our experiments). While the definition of trust varies widely in the literature, a common definition is “a single, global, unidimensional measure of trust” (Geyskens et al., 1998, p. 225). We include a global measure: “Overall, this company is trustworthy.” Other authors include benevolence and credibility (e.g., Doney & Cannon, 1997; Ganesan, 1994; Ganesan & Hess, 1997; Geyskens et al., 1998). We implemented benevolence with “I believe that this company is willing to assist and support me,” and we implemented credible with “Overall, this company has the ability to meet customer needs.” Sirdeshmukn et al. (2002) suggest that trust includes operational competence, which we implemented with “The company is very competent in its dealings with its customers.” Finally, Anderson and Narus (1990, p. 45) suggest that the trustworthy firm “will perform actions that will result in positive outcomes for the [consumer]” and Anderson and Weitz (1988) define trust as a [consumer’s] belief that “its needs will be fulfilled in the future by actions undertaken by the other party.” We implemented positive outcomes with “This company makes excellent vehicles.” All five items were measured with 7-point Likert questions. Consistency with one’s products, listening to customers and placing customers ahead of profit are becoming increasingly important to consumers, as indicated by the Edelman Trust Barometer (http://trust.edelman.com/what-the-2012-edelman-trust-barometer-means-for-purpose/).

Together, the five items reflected the goals of AAM on how they might engender trust among consumers. Before we use the five items to evaluate the impact of competitive information, we establish that the five items form a unidimensional construct and that the construct is related to overall trust. Using methods recommended in Churchill (1979), we purify the scale. The scale is unidimensional with high construct reliability and maximized with all five items: Cronbach’s α = 0.95. As a further test, we compare a four-item scale without the global item; the reduced scale is highly correlated with the global item (ρ = 0.88). We therefore conclude that the five items represent a single scale and that it is reasonable to call the scale “trust” for the purpose of evaluating the implications of competitive information. In the following analyses, trust is measured by the sum—score (divided by the number of items such that the trust scale ranges from 1 to 7).

### 3.5. Randomization tests in year 1

All models of consideration and purchase use treatment-assignment dummies. Following standard experimental procedures, the impact of a treatment is measured for all respondents for whom the treatment was available whether they experienced the treatment. This approach is conservative and avoids effects that might be due to differential take-up of the treatments. For completeness, we compared treatment-assignment analyses to self-reported-treatment analyses. The pattern of coefficients and their significance was similar for both analyses.

Qualitative data are consistent with the hypothesis that take-up was random in year 1. For example, a portion of consumers experienced
technical difficulties with the competitive online advisor, and a few could not come to the competitive test drive due to last-minute scheduling issues. Take-up rates were 91.1% for competitive test drives, 99.4% for brochures, 97.4% for the community forum, and 82.1% for the online advisor. The high take-up rates and the similarity of coefficients suggest that treatment take-up (given it was offered) was sufficiently random that take-up selection had little or no effect in year 1. We thus leave analyses of self-selected take-up to year 2 when the trust signal is more cleanly implemented.

Although we use treatment assignments as independent measures, it is useful to examine further whether take-up was random. Specifically, we examine consumers who (1) were not assigned to a treatment, (2) were assigned to a treatment but did not report participation, and (3) were assigned and reported participation. If there were non-random take-up, then consumers in group (3) would be a non-random draw from (2 and 3). However, if that were true, non-random take-up from (2 and 3) to (3) would leave a non-random set of consumer behaviors in (2). We find no differences in the dependent measures between groups (1) and (2), thus providing further evidence that take-up was sufficiently random. For example, measured brand consideration does not vary between groups (1) and (2) for competitive test drives (t = 3.6, p < .01). The treatment vs. control difference is not significant for customized brochures, the competitive forum, and the competitive advisor. The increase in cumulative purchase follows the same pattern with an 11.1% difference for competitive test drives (t = 2.4, p = .02) and insignificant effects for the other treatments. The impact on trust is similar but not identical. Competitive test drives increase trust (11.6%), but the competitive advisor has a negative effect (−8.9%), and both are significant at p < .01. Neither customized brochures nor the competitive forum change trust significantly (we examine more-complete models and trust mediation in Section 5).

We might posit the effect of competitive information to be either larger or smaller among consumers who own AAM vehicles. This effect might be larger because current vehicles are improved relative to prior vehicles, but it might be smaller because consumers who do...
not own AAM vehicles have less experience with older, less well-received vehicles. Empirically, the data suggest that there is no interaction effect due to prior AAM ownership, implying either that the two effects cancel or that neither is strong. Competitive test drives is the only treatment with a significant impact among non-AAM owners, and the magnitude of that impact is virtually identical to the magnitude among all consumers (lower left of Table 2). We explore interactions more formally in Section 5. We find no differential impact due to age or sex. The age–sex comparisons are not shown to simplify the tables.

Analyses of the main effects are the simplest and cleanest set of analyses. Heterogeneity is mitigated because of randomization in treatment assignment. However, we can improve insight by accounting for dynamics, interactions among treatments, the conditioning of purchase on brand consideration, the potential for trust mediation, and heterogeneity.

5. Dynamics of trust, brand consideration, and purchase

5.1. Average trust, brand consideration, and purchase by month

Table 3 summarizes average trust, brand consideration, and purchase by treatment for months 2 to 6. The response to competitive information is more complex. For example, trust increases substantially in month 4 among consumers who experience test drives and declines in months 5 and 6, but it does not decline to pre-month 4 levels. A large change in brand consideration occurs in month 4 when consumers experience competitive test drives, but the effect endured through subsequent months. These observations might be due to an effect where competitive test drives increase trust which then decays slowly rather than immediately. In addition, in automotive markets, a consumer might come to a competitive test drive in one month, consider and seriously evaluate vehicles in the next month, and purchase in a third month. To untangle these effects, we need analyses beyond mere inspection of Table 3. We need analyses based on a theory of automotive purchasing dynamics.

Table 3 reports the main effects, but the experiment is a fully crossed $2 \times 2 \times 2 \times 2$ design with some consumers experiencing other forms of competitive information in earlier months, some in later months, and some not at all. More sophisticated analyses are necessary to account for the fully crossed design, potential interactions among treatments, and the time-varying nature of the treatments. Finally, the measures in Table 3 are averages and do not account for individual differences. To untangle the true effect of competitive information we now model the complexities of the dynamics of the automotive market, the complexity of the experimental design, and potential individual differences.

5.2. Trust dynamics

Trust builds or declines over time. A consumer might experience a treatment in month $t$ and, as a result, increase his or her trust in a brand. However, the consumer may not trust the brand enough to consider it. Another treatment in month $t+1$ might increase trust further and be enough to encourage brand consideration. To capture this phenomenon, we model trust as a variable that can increase or decrease over time as a result of treatments. Trust might also decay. Specifically, we model trust in month $t$ as a sum of $\gamma$ times trust in month $t-1$ plus effects due to the treatments in month $t$, where $\gamma$ and the coefficients of the treatment effects are to be estimated ($\gamma \leq 1$). We provide detailed equations in Section 6.

5.3. Interactions among the treatments

Hauser, Urban, and Weinberg (1993) examine consumers’ search for information concerning automobiles and find that the value of an information source sometimes depends upon whether consumers had previously experienced another information source. These researchers’ data suggest two-level interactions but no three-level interactions. In light of this prior research, we examine models that allow interactions among the treatments.

5.4. Individual differences among consumers

Although the treatments were randomized, consumers with different purchase histories and different demographics (age and sex) might react differently. For example, consumers who now own AAM vehicles may base their trust, brand consideration, or purchase on their prior ownership experience. To capture purchase history,
we include dummy variables for “own other American” and “own Japanese.” There is no dummy variable for “own European.”

Heterogeneity might also be unobservable. One way to correct for unobserved heterogeneity in the propensity to trust, consider, or purchase an AAM vehicle would be to compute month-to-month differences in trust, brand consideration, and purchase. However, both brand consideration and purchase are quantal (0 vs. 1) measures and month-to-month differences would implicitly assume (a) no decay and (b) perfect reliability of repeated measures. With repeated noisy measures, the best estimate of the true score at t is not the score at t−1 but rather a function of reliability times the lagged score (Nunnally & Bernstein, 1994, 222). While we can never rule out unobserved heterogeneity completely, we can examine whether unobserved heterogeneity is likely to be a major effect or a second-order effect. Specifically, (a) a coefficient close to one in a trust regression (γ = 0.833, yet to be shown), (b) explicit controls for observable heterogeneity, (c) consistency with the main-effect analyses (yet to be shown), and (d) continuous-time Markov analyses based on differences (yet to be shown) all suggest that effects due to unobserved heterogeneity are negligible and unlikely to reverse primary insights.

6. Modeling dynamics in year 1

Ignoring for a moment dynamics, persistence, more complete prior-ownership effects, interactions among treatments, and unobserved external shocks, we see that trust is correlated with both brand consideration (ρ = 0.22) and purchase (ρ = 0.17), both significant at the 0.01 level. The correlation with lagged trust is higher for brand consideration (ρ = 0.61) but lower for purchase (ρ = 0.5). This result is consistent with a dynamic interpretation that trust at the beginning of a month is the driver of brand consideration during that month. However, these simple correlations do not capture all of the dynamics and, technically, misuse a correlation coefficient for two quantal outcomes (brand consideration and purchase).

To account for theoretical dynamics and to respect the quantal nature of the outcome variables, we use conditional-logit analyses of brand consideration and purchase (see Fig. 2). Specifically, we ask whether the treatments increase brand consideration and, among those consumers who consider AAM vehicles, whether the treatments also affect purchase.

In the conditional-logit analyses, we include lagged brand consideration as an explanatory variable to focus on changes in brand consideration. We include dummy variables for observation months to account for unobserved marketing actions and environmental shocks (month 1 is a pre-measure and the month-2 dummy variable is set to zero for identification). The month dummy variables also account for any measurement artifact that might boost brand consideration (e.g., “Hawthorne” effect). To account for observed heterogeneity in past purchases, we include prior ownership of AAM, other American, and Japanese (relative to European) vehicles. Age and sex effects were examined but suppressed to simplify Table 4, as they were not significant.

Let \( R_t \) be a measure of consumer i’s trust in month t and let \( y_{it0} = 1 \) if consumer i was assigned to treatment j in month t, and \( y_{it0} = 0 \) otherwise. Let \( y_{it1} = 1 \) if consumer i has characteristic k, and let \( y_{itk} = 0 \) otherwise. Let \( \delta_t = 1 \) in month t and \( \delta_t = 0 \) otherwise. Trust dynamics are modeled with Eq. (1) where we estimate the \( \gamma, \rho, \delta, u, b \):

\[
R_t = \gamma R_{t-1} + \sum_{j=1}^{4} w_j y_{ij} + \sum_{k=1}^{K} v_k y_{ik} + \sum_{t=3}^{6} u_t \delta_t + b^R. 
\]

Let \( C_{it} = 1 \) if consumer i considers an AAM brand in month t and \( C_{it} = 0 \) otherwise. Let \( P_{it} = 1 \) if consumer i purchases an AAM vehicle in month t and \( P_{it} = 0 \) otherwise. The conditional-logit models are specified by Eq. (2):

\[
\text{Prob}(C_{it} = 1) = \frac{e^{\gamma R_{t-1} + \sum_{j=1}^{4} w_j y_{ij} + \sum_{k=1}^{K} v_k y_{ik} + \sum_{t=3}^{6} u_t \delta_t + b^C}}{1 + e^{\gamma R_{t-1} + \sum_{j=1}^{4} w_j y_{ij} + \sum_{k=1}^{K} v_k y_{ik} + \sum_{t=3}^{6} u_t \delta_t + b^C}}.
\]

\[
\text{Prob}(P_{it} = 1|C_{it} = 1) = \frac{e^{\delta R_{t-1} + \sum_{j=1}^{4} w_j y_{ij} + \sum_{k=1}^{K} v_k y_{ik} + \sum_{t=3}^{6} u_t \delta_t + b^P}}{1 + e^{\delta R_{t-1} + \sum_{j=1}^{4} w_j y_{ij} + \sum_{k=1}^{K} v_k y_{ik} + \sum_{t=3}^{6} u_t \delta_t + b^P}}.
\]

6.1. Direct effects of treatments

We begin with main effects of the treatments, as shown in the first and second columns of parameters in Table 4. The brand consideration analysis includes all respondents. The purchase analysis is conditional on brand consideration: only those respondents who consider AAM in the month of measurement on purchase. The purchase model explains 24.8% (\( U^2 \), sometimes called a pseudo-\( R^2 \)), measures the percent of uncertainty explained, Hauser, 1978). Brand consideration is increased if consumers own AAM or other American vehicles and decreased if they own Japanese vehicles. Brand consideration is also higher in months 3 to 6 relative to month 2. The only significant direct treatment effect is due to competitive test drives. Purchase, conditional on brand consideration, also increases with competitive test drives (marginally significant), but there are no direct effects of prior ownership or month of measurement on purchase. The purchase model explains less uncertainty, suggesting that the treatments affected brand consideration more strongly than purchase.

6.2. Trust as a mediator

There is ample precedent in the literature for trust as a mediator of purchase or purchase intentions (e.g., Bart et al., 2005; Büttner & Görnitz, 2008; Erdem & Swait, 2004; Morgan & Hunt, 1994; Porter & Donthu, 2008; Urban et al., 2009; Yoon, 2002). In a series of experiments, Trifts and Häubl (2003) demonstrate that competitive price information affects preference, but the effect on preference is mediated through trust.

We use the methods of Baron and Kenney (1986) to test whether competitive information treatments were mediated through trust. Specifically, if the treatments affect trust and if the treatments affect brand consideration (or purchase), we estimate a third model. We add an indicator of trust as an explanatory variable in the conditional-logit models. If the treatments are mediated through trust, then (1) the
indicator of trust should be significant in the new models, and (2) the direct effect of treatments on consideration (or purchase) should now be insignificant. Partial mediation includes (1) but requires only that the direct effect decrease in magnitude.

We must be careful when we add trust to the model. We use lagged trust to be consistent with the dynamics of measurement and causality. Lagged trust has the added benefit that joint causality in measurement errors is reduced because the trust measures occur in different months than the brand consideration and purchase measures. Nonetheless, to account for unobserved shocks that affect trust in month t − 1 and brand consideration (purchase) in month t, we use estimated lagged trust in an equation that predicts brand consideration (purchase) with reported treatments [see Online Appendix for other trust regressions]. That is, we add $R_{t−1}$ as an explanatory variable on the right-hand sides of Eq. (2) where $R_{t−1}$ is estimated with Eq. (1). Traditional mediation analyses use lagged trust directly. In our data, these tests also indicate mediation and have similar implications.

We first examine the trust regression. Competitive test drives clearly increase trust, and there is evidence that customized brochures increase trust. The impact of customized brochures is consistent with published studies of customization (e.g., Ansari & Mela, 2003; Hauser et al., 2010). The effect of customized brochures is less apparent in the main-effect analyses because, although the effect is strong in earlier months, it becomes insignificant in the last month. Review Table 3. Consistent with the main-effect analyses, the conditional-logit analyses and the trust regression identify no impact on brand consideration and purchase for the community forum and the competitive advisor.

We now add lagged estimated trust to the conditional-logit analyses. Such two-stage estimates are limited-information maximum-likelihood estimates. The two-stage estimates are consistent but require bootstrap methods to estimate the standard errors for the coefficients (Berndt, Hall, Hall, & Hausman, 1974; Efron & Tibshirani, 1994; Wooldridge, 2002, p. 354, 414). The parameter estimates and standard errors are based on 1000 bootstrap replications (Table 4 reports significance; standard errors are available upon request).

Following Baron and Kenny (1986), the treatments are mediated through trust if (a) including lagged trust in the model increases fit significantly (and the lagged trust variable is significant) and (b) the treatments are no longer significant when estimated lagged trust is in the model. The increase is significant for brand consideration and marginal significant for purchase ($\chi^2 = 86.7, p < .001$ and $\chi^2 = 3.1, p = 0.08$, respectively). Once we partial out lagged trust, there remain no significant direct effects due to the treatments. This result suggests trust mediation.

6.3. Testing interaction effects for prior ownership and for multiple treatments

Prior ownership might influence the impact of the treatments and there might be interactions due to multiple treatments. To test whether prior ownership affects the impact of competitive information, we cross-prior ownership of an AAM vehicle with the treatment-assignment dummies. For trust, brand consideration, and purchase, the interactions are not significant ($F = 1.91, p = .11$; $\chi^2 = 4.3, p = .37$, and $\chi^2 = 7.0, p = .13$, respectively).

We also tested interactions among the treatments. Treatment interaction-effects do not add significantly to a trust regression using a specification that allows all interactions ($F = .85, p = .59$). We continue to use estimated lagged trust (without interactions) and estimate a conditional-logit model that allows interactions. The fully saturated brand consideration model that allows all possible interactions is not significant relative to a main-effects model and provides no additional insight ($\chi^2 = 12.0, p = .10$). A few coefficients are significant, but all include competitive test drives with slight variations in parameter magnitudes that depend upon the other combinations of treatments. To avoid over-fitting with a fully saturated model, we examined a more parsimonious model in which we add a variable for two or more treatments. This parsimonious model is consistent with earlier automotive studies (see, e.g., Hauser et al., 1993). The “two or more treatments” variable is not significant for brand consideration or purchase. Neither the fully saturated nor the parsimonious analysis highlights any interpretable interactions suggesting that the fully saturated model was over parameterized. The fourth (brand consideration) and sixth (conditional purchase) data columns of Table 4 display models with

---

Table 4 Conditional-logit analyses and trust regression with time-specific stimuli — year 1 random assignments.

<table>
<thead>
<tr>
<th>Dependent measure</th>
<th>Direct effects not mediated</th>
<th>Mediated by trust (bootstrap estimates)</th>
<th>Trust regression (lagged trust is used in this regression)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consider brand</td>
<td>Purchase if consider</td>
<td>Consider brand</td>
</tr>
<tr>
<td>Constant</td>
<td>−1.492</td>
<td>−2.567</td>
<td>−3.945</td>
</tr>
<tr>
<td>Lagged brand consider</td>
<td>2.537</td>
<td>2.394</td>
<td>2.405</td>
</tr>
<tr>
<td>Lagged trust bat</td>
<td>.531</td>
<td>.392</td>
<td>.346</td>
</tr>
<tr>
<td>Competitive test drives</td>
<td>.579</td>
<td>.938</td>
<td>.392</td>
</tr>
<tr>
<td>Customized brochures</td>
<td>.079</td>
<td>.477</td>
<td>.059</td>
</tr>
<tr>
<td>Competitive forum</td>
<td>.144</td>
<td>.122</td>
<td>.133</td>
</tr>
<tr>
<td>Competitive advisor</td>
<td>−.023</td>
<td>−.103</td>
<td>.136</td>
</tr>
<tr>
<td>Prior ownership AAM</td>
<td>.390</td>
<td>.137</td>
<td>.327</td>
</tr>
<tr>
<td>Prior own other American</td>
<td>.304</td>
<td>−.005</td>
<td>.253</td>
</tr>
<tr>
<td>Prior ownership Japanese</td>
<td>−.577</td>
<td>−.188</td>
<td>−.464</td>
</tr>
<tr>
<td>Month 3</td>
<td>.313</td>
<td>.200</td>
<td>.461</td>
</tr>
<tr>
<td>Month 4</td>
<td>.419</td>
<td>.264</td>
<td>.433</td>
</tr>
<tr>
<td>Month 5</td>
<td>.523</td>
<td>−.238</td>
<td>.396</td>
</tr>
<tr>
<td>Month 6</td>
<td>.722</td>
<td>.185</td>
<td>.654</td>
</tr>
<tr>
<td>Prior ownership of AAM crossed with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive test drives</td>
<td></td>
<td></td>
<td>−.211</td>
</tr>
<tr>
<td>Customized brochures</td>
<td></td>
<td></td>
<td>.109</td>
</tr>
<tr>
<td>Competitive forum</td>
<td></td>
<td></td>
<td>−.405</td>
</tr>
<tr>
<td>Competitive advisor</td>
<td></td>
<td></td>
<td>−.444</td>
</tr>
<tr>
<td>Two or more treatments</td>
<td></td>
<td></td>
<td>.208</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−820.6</td>
<td>−218.2</td>
<td>−777.2</td>
</tr>
<tr>
<td>$U^2$ (aka pseudo-$R^2$)</td>
<td>24.8%</td>
<td>3.1%</td>
<td>28.8%</td>
</tr>
</tbody>
</table>

Sex and age coefficients are not shown (not significant). Trust regression interactions are not significant.

a Significant at the 0.05 level.
b Significant at the 0.10 level.
interactions due to prior ownership and due to two or more treatments. The addition of prior ownership and interactions among the treatments are not significant ($\chi^2 = 4.9, p = .42$ and $\chi^2 = 7.1, p = .21$ for brand consideration and purchase, respectively).

6.4. Interpretation of year-1 results relative to year-1 hypotheses

**H4** is clearly supported. Enhanced trust at the end month $t - 1$ is significantly correlated with brand consideration in month $t$. When there is an effect due to competitive information, it is mediated through trust. Enhanced trust at the end of month $t - 1$ has a significant effect on consideration in month $t$ and a (marginally) significant effect on purchase in month $t$. These field-experiment results are consistent with results from the structural-equation analysis of questionnaires and laboratory experiments found in the literature. More importantly, given the in vivo nature of the field experiment and the modeling of dynamics, causality and external validity are likely.

**H1** through **H3** are more complicated. Contrary to predictions and recommendations in the literature, all forms of competitive information do not enhance trust. Neither the competitive forum (word of mouth) nor the competitive advisor enhanced trust, brand consideration, or purchase. Only competitive test drives and possibly customized brochures enhanced trust, brand consideration, and purchase. Not only was this complexity not predicted by theory, but the results also surprised AAM. All four generic forms of competitive information were provided altruistically and made AAM vulnerable, but vulnerability and altruism were not sufficient to affect trust, and through trust, brand consideration and purchase.

The experiments alone do not explain why some forms of competitive information were better than others, but our belief and the judgment of managers at AAM suggest that the valence of the information made a difference. AAM does well relative to competition in competitive test drives; the customized brochures highlighted the benefits of AAM. However, there was substantial negative information in both the competitive forum and the competitive advisor. Empirically, it appears that the effect of negative information countered the effect of trust from competitive information. We believe that this possibility is the most likely interpretation, but we cannot rule out other interpretations, such as the strength of the signal. We can rule out cheap talk because both the most costly and least costly forms of competitive information were effective. This moderation of the vulnerability/altruism-to-trust link by the valence of information is worthy of further analyses.

Although it may seem intuitive a posteriori, an interpretation that negatively valenced information does not enhance trust refines theories of trust signaling. Many authors posit only that the firm should act altruistically and provide information that is truthful and dependable (e.g., Erdem & Swait, 2004; Urban et al., 2009; Urban et al., 2000). Prior to the field experiments, the managers at AAM believed that vulnerability and altruism would signal trust and that the enhanced trust would lead to brand consideration. The insight about the valence of the competitive information was managerially important.

6.5. Continuous-time analyses

As a final check, we estimate continuous-time Markov models that relax three restrictions. These Markov models allow flows to happen in continuous time (even though observations of the results of those flows are at discrete time intervals). In addition, because the dependent variables are differences in observed states (not consider, consider but not purchase, and consider and purchase), the Markov models are analogous to difference-in-differences models and thus account for unobserved heterogeneity in the propensity to consider AAM. Finally, all transitions are estimated simultaneously with a single likelihood function (details are in Appendix A). The results reinforce the implications of the conditional-logit analyses: competitive test drives have a significant effect on brand consideration, that effect is mediated through trust, and neither competitive forums nor competitive advisors have significant effects.

7. Year 2: field test of trust signaling

Year 1 established that positively valenced competitive information (competitive test drives and, possibly, customized brochures) enhances trust, which, in turn, precedes brand consideration. However, the most effective competitive information (competitive test drives) is extremely expensive to provide and may not be cost-effective. Conversely, competitive information might be cost-effective if a firm could signal trust-worthiness by simply offering to make competitive information available. If the signal alone were sufficient, a national launch would be feasible. Trust signaling (H5 through H8) suggests that the offer of competitive information itself engenders trust. However, there is a danger that a trust signal would attract only those consumers who already trust and/or feel favorably toward the brand that is signaling trust. If this possibility is supported, H5 through H8 might be rejected.

To test trust signaling, year-2 consumers were assigned randomly to one of two groups. Consumers in the control group received no treatments whatsoever. Consumers in the test group received an advertisement inviting them visit a “My Auto Advocate” website (screenshots available in an Online Appendix). We call consumers who visited the website based on the advertisement alone the “website-visited” test group. Consumers in the test group who did not visit the “My Auto Advocate” website in response to advertising were invited to an “Internet study” that included a visit to the “My Auto Advocate” website. We call these consumers the “website-forced-exposure” test group. At the “My Auto Advocate” website, both the website-visited test group and the website-forced-exposure test group could select (opt in to) any combination of five treatments. Taken together, these two sub-groups make up the test group that tests trust signaling. The website-forced-exposure group represents a stronger signal and more intensive national advertising.

The panel ran monthly from January to June 2005. In year 2, members of Harris Interactive’s panel were screened to be in the market for a new vehicle, on average within the next 2.2 years, and invited to participate and complete six monthly questionnaires. This 2.2-year average was designed to draw in more consumers (relative to the year-1 twelve-month intenders). Once consumers visited the “My Auto Advocate” website, they were given incentives to opt in to the treatments, but unlike in year 1, they were not assigned to treatments. For example, consumers received 20 reward certificates (worth $1 each) for participating in the competitive test drives. Incentives for the other treatments were the order of 5 reward certificates.

Harris Interactive invited 6092 Los Angeles consumers of which 1720 completed all six questionnaires for an average response/completion/retention rate of 21.7%. This rate was not significantly different across the three groups (control vs. website-visited vs. website-forced-exposure, $p = .25$). Brand consideration, purchase, and trust were measured as in year 1.

7.1. Treatments in year 2

The treatments in year 2 were similar to year 1. The competitive test drive treatment and the word-of-mouth treatment were virtually the same with minor updates. The competitive online advisor was improved slightly with a better interface and a “garage” at which consumers could store vehicle descriptions. The online advisor still favored other manufacturers’ vehicles in year 2, although somewhat less so than in year 1. The major change was the brochures, Year 2 used electronic brochures for AAM vehicles (called eBooks). These brochures were online or downloadable, not mailed, and were less customized. An additional treatment, eBrochures, allowed consumers to download...
competitive brochures. Although many competitive brochures were available on automakers’ websites, the single-source webpage made it more convenient for consumers to compare vehicles. Screenshots for “My Auto Advocate” and the treatments are available in an Online Appendix. Table 5 summarizes the numbers of consumers who opted in to treatments in year 2. All treatments except competitive test drives were reasonably popular and available in all months.

7.2. Testing whether the trust signal was effective

We first examine brand consideration and purchase in the test (website-visited and website-forced-exposure) vs. the control groups. There were no significant differences in trust, brand consideration, or purchase intentions (t = .9, p = .37, t = .14, p = .88, and t = .18, p = .86, respectively). Similarly, the differences between the website-visited sub-group and control group were not significant (t = 1.4, p = .15, t = −1.5, p = .14, and t = −.05, p = .96, respectively). These results suggest that a trust signal provided little or no lift in trust, brand consideration, and purchase relative to the control. Contrary to extant theory, we reject H5, H6, and H7. Because these hypotheses are rejected, we cannot test the conditional hypothesis, H8.

There are at least two complementary explanations for the null effect in year 2. First, the null effect may be because offering competitive information does not increase trust. In this case, it might have been that only those consumers who already trusted AAM were likely to opt in to the treatments. A second (and complementary) explanation is that fewer consumers opted in to the effective treatments (test drives and possibly brochures) than potentially negative treatments (word-of-mouth and online advisors).

7.3. Examining potential explanations

If the signal did not enhance trust, brand consideration, and purchase in the test group relative to the control group, then we should see evidence that consumers who were more favorable to AAM opted in to the treatments (to be most effective, the signal needed to reach consumers who do not trust the automaker, not those who already trust the automaker). To test whether the trust signal reached targeted consumers we compare consumers in the control group (who received no treatments) to those in the test group who did not opt in to any treatments.

Among no-treatment consumers, the non-treated members of the test group had significantly lower trust, brand consideration, and purchase intentions than the control group (t = 6.1, p = .0, t = 6.1, p = .0, and t = 2.0, p = .05, respectively). By implication, consumers who opted in to competitive-information treatments were consumers who were more trusting of AAM (or, at least, more favorable toward the automaker). Comparing the control group to non-treated members of the website-forced-exposure test group gives similar results. The trust signal ultimately targeted consumers more likely to be favorable toward AAM. The signal alone did not engender trust, brand consideration, and purchase.

We gain further insight by redoing for year 2 the main-effects analyses (as in Table 2) and conditional-logit analyses (as in Table 4). Details of the year-2 analyses are available in an Online Appendix. The opt-in main effects (year 2), relative to random-assignment main effects (year 1), are consistent with a hypothesis that for each treatment, consumers who opted in to that treatment were a priori more favorable to AAM. When we account for prior ownership, prior propensities, dynamics, and mediation with conditional-logit analyses, we find that the effects of the treatments are consistent with those observed in year 1. Thus, the null effect of the trust signal is probably observed because the signal did not encourage opt in among those to whom the signal was targeted.

8. Hypotheses revisited

Table 6 summarizes the implications of the two field experiments. H1 through H4 suggest that if firms build trust with consumers, trust will lead to brand consideration and purchase. H5 through H8 suggest that trust signals alone should achieve these outcomes. The situation in the automotive industry in 2003–2005 provided an excellent test of these theories. Because of past experiences with AAM’s vehicles, many consumers would not even consider those vehicles in 2003–2005. Because the vehicles from AAM had improved relative to prior years, the automaker had an opportunity to build trust by providing competitive information (year 1) or by signaling trustworthiness (year 2). Current theories suggest altruism and vulnerability will build trust but do not distinguish whether the information provided to the consumer should favor the brand. In fact, altruism and vulnerability are greater if the competitive information does not always favor the firm’s brands.

The year-1 experiments were consistent with H4. Trust in one month was correlated with brand consideration in the next month, and the effect of competitive information was mediated through trust. However, the year-1 experiments also identified certain competitive information types as more effective than other competitive information types. Experiential information (competitive test drives) was the most effective communications strategy. Tangible experience convinced consumers that the automaker’s products had improved relative to the competition.

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers who selected treatments in year 2 signaling experiment (test of signaling trust through advertising then website opt in to treatments).</td>
</tr>
<tr>
<td>Number of respondents who selected the indicated treatment in that month.</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Competitive test drives</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Competitive eBrochures</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>AAM eBooks</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Competitive forum</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Competitive advisor</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

⁴ Website-visited = consumers visited the “My Auto Advocate” website on their own and opted in to the treatment.
⁵ Website-forced = consumers were forced to visit the “My Auto Advocate” website, but opted in to the treatment.
⁶ Not treated = consumers in control group plus consumers in test group who did not opt in to the treatment.
Table 6
Summary of hypotheses, support, tests, and implications.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Support or not supported</th>
<th>Evidence</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1. Trust enhanced by competitive information (in test vs. control).</td>
<td>Supported for positively valenced information but not for information with substantial negative content (positive vs. negative remains a hypothesis).</td>
<td>Competitive test drives and customized brochures have significant effects in trust regressions. Other treatments have no significant effect.</td>
<td>There exist situations where competitive information leads to trust, especially positively valenced information. Negatively valenced information may not engender trust. If a firm can engender trust it can enhance brand consideration.</td>
</tr>
<tr>
<td>H2. Brand consideration enhanced by competitive information (in test vs. control).</td>
<td>Strongly supported for positively valenced information.</td>
<td>Trust-to-brand consideration is significant in conditional-logit and continuous-Markov analyses.</td>
<td>If a firm can engender trust it might be able to obtain sales.</td>
</tr>
<tr>
<td>H3. Sales enhanced by competitive information (in test vs. control).</td>
<td>Marginally supported for positively valenced information.</td>
<td>Trust-to-sales is marginally significant in conditional-logit and significant in continuous-Markov analyses.</td>
<td>Other means to gain trust might gain brand consideration and sales.</td>
</tr>
<tr>
<td>H4. Mediation by trust.</td>
<td>Strongly supported.</td>
<td>Baron–Kenney tests support trust mediation in both conditional-logit and continuous-Markov analyses.</td>
<td>Signaling altruism (trust) by providing competitive information did not enhance trust, consideration, or sales. Consumers must experience the information.</td>
</tr>
<tr>
<td>H5–H7. Offering to provide competitive information enhances trust, brand consideration, and sales.</td>
<td>Rejected. Opt-in communications attracted consumers who are already favorable.</td>
<td>Year–2 null effects. Also comparisons of control consumers to consumers in the test groups who did not opt in to treatments.</td>
<td>H8 is conditioned on support for H5–H7.</td>
</tr>
</tbody>
</table>

Neither word-of-mouth (community forums) nor competitive advisors increased trust in the test group relative to the control group. From the year-1 experiments alone, we cannot determine whether these forms of competitive information are not effective or whether AAM’s implementation of these forms of competitive information was not effective. Community forums relied on other consumers’ opinions — opinions contaminated with past experience. Online advisors relied in part on past consumer experience and may have lagged improvement in vehicles. We posit that positively valenced information enhanced and negatively valenced information diminished the effectiveness of competitive-information implementations.

Hypothesis. Unbiased competitive information can build trust, and trust enhances brand consideration and purchase. The firm builds trust if competitive information enables the firm to communicate to consumers that it is acting altruistically and that it has products that meet their needs. However, the availability of competitive information alone does not enhance trust; consumers must process the competitive information.

We believe that this hypothesis applies across a variety of situations and product categories as discussed in the introduction to this paper. Naturally, this hypothesis is subject to tests in different categories with different implementations of generic competitive-information treatments and with different signals of trust.

9. Cost effectiveness and managerial implications

We now move from theory back to practice. Because the experiments established that some forms of competitive information engender trust and that trust leads to brand consideration, AAM ultimately implemented competitive-information strategies. The automaker first used the more general insights to refine competitive test drives. The following calculations illustrate the motivation behind managers’ decisions at the time. We disguise the proprietary data with comparable publicly available data.

For illustration, we assume a 15% market share. Based on this share, competitive test drives provide an 11.1% sales lift (year-1 data) with an approximate cost of $120 per participating consumer (with incentives). When the financial situation improved, AAM tested competitive test drives for SUVs with a dealer in Arizona (automotive consumers often use agendas to screen on body-type; Hauser, 1986; Urban, Librali, MacDonald, Bordley, & Hauser, 2012). These competitive test drives proved to be cost-effective — approximately $100–200 per incremental sale. Costs were substantially lower because the test drives were from the dealer’s lot (no need to rent a test track), because fewer vehicles were necessary (only SUVs), and because the dealer could borrow or rent vehicles from competitive dealers. On the benefit side, gross margins were higher than average for SUVs. AAM continued to experiment with competitive test drives in key local markets when high-value skeptical consumers could be targeted cost-effectively. In late 2010, the head of U.S. marketing for AAM launched a yearlong series of weekend competitive test drives at dealerships. Each event invited a few thousand potential buyers to compare AAM vehicles with competitive vehicles.

Year 2 taught managers that they needed to do more than signal that competitive information was available. Communication strategies should be targeted at consumers who do not already trust the automaker. Managers now place a premium on targeting competitive information toward skeptical consumers. New methods include interactive screening to identify consumers who answer questions that indicate they do not trust AAM. Targeted consumers would receive substantial incentives to participate in competitive-information treatments.

Based on the 2003–2005 data and managerial judgment, managers believe that providing competitive information is effective when there is good news and when it is cost effective. As of this writing, AAM includes key competitive comparisons on its website using standardized Polk data on prices, specifications, and equipment for preselected competitive and consumer-specified vehicles. In 2009, AAM used a national advertising campaign that encouraged consumers to compare AAM vehicles with those of competitors regarding fuel economy and styling. Many AAM dealers offer unsolicited extended weekend test drives and encourage competitive comparisons. AAM believes that competitive information builds trust, brand consideration, and sales and is profitable only if implemented cost-effectively to skeptical consumers for categories in which AAM has good vehicles relative to competitors. This more nuanced trust-based strategy is believed to be more profitable than a general strategy of trust signaling.
Online Appendices available from the authors

OA1. Screenshots of year-2 treatments.
OA3. Main effects, trust regression, and conditional-logit analyses in year 2.

Year 1 and year 2 data as used in tables are available from the authors.

Appendix A. Continuous-time Markov analysis in year 1

The analyses in the text model month-to-month dynamics but do not allow flows to happen in continuous time nor do they allow reverse flows from “consider” to “not consider.” We address these issues with continuous-time Markov analyses (Cox & Miller, 1965; Hauser & Wisniewski, 1982, hereafter “Markov” analyses). There are two added advantages of Markov analyses: (a) a single likelihood function estimates all parameters for all defined flows simultaneously, and (b) treatments affect differences in behavioral states directly. By the Markov property, observing a customer in “consider” in month t is treated differently if the customer was in “not consider” versus in “consider” at month t – 1. By focusing on differences in behavioral states, the Markov analyses are less sensitive to unobserved heterogeneity.

The Markov analyses complement the conditional-logit analyses in Fig. 2; the concepts are similar, but we model a more complete set of flows and allow the flows to occur in continuous time (Fig. A1). Consumers’ “flow” among states with instantaneous flow rates dependent upon the treatments and other variables. Mathematically, for \( j \neq i \), \( a_{ij} \Delta t \) is the probability that the consumer flows from state i to state j in the time period between t and \( t + \Delta t \) for very small \( \Delta t \) during the nth month. We specify as relevant the flow rate as a log-linear function of the treatment assignments, prior ownership, age, sex, month dummies, and interactions as relevant — the same types of specifications as in the conditional-logit analyses. Although we model instantaneous flow rates, we only observe the state that describes each consumer at the end of each month. Fortunately, using the \( a_{ij} \)'s, we can calculate the probability, \( p_{ijn} \), that the consumer was in state i at the beginning of the nth month and in state j at the end of the month. Specifically,

\[
P_{ij n} = e^{A n (T_{n} - T_{n-1})} \equiv \sum_{t=0}^{\infty} \frac{A_{i}^{t} (T_{n} - T_{n-1})^{t}}{t!} \equiv V_{n}[\exp A_{n}] V_{n}^{-1},
\]

where \( P_{ij} \) is the matrix of the \( p_{ij} \)'s, \( A_{n} \) is the matrix of the \( a_{ij} \)'s, \( T_{n} \) is the time at the end of the nth month, \( V_{n} \) is the matrix of eigenvectors of \( A_{n}(T_{n} - T_{n-1}) \), and \( [\exp A_{n}] \) is the matrix with the exponentiation of the eigenvalues on the diagonal.

Prior applications in marketing used regression approximations of Eq. (A1) (Hauser & Wisniewski, 1982). With today’s computers, we use maximum-likelihood methods with all flows estimated simultaneously. See Kulkarni (1995) for a review of computational methods to deal with matrix exponentiation. While we would like to repeat the Markov analyses for all of the specifications tested by conditional-logit analyses, the convergence of the Markov estimates and the computation times appear to be most appropriate for more parsimonious models. Thus, we use the Markov analyses as a confirmation of the conditional-logit analyses by carefully selecting the explanatory variables based on the conditional-logit analyses (we do not need lagged consideration in the Markov analyses because the analyses are based on transitions from “not consider”, rather than based on estimating consideration as a function of lagged consideration and other variables). For simplicity of exposition, we report key analyses in Table A1. Other analyses and R-code are available from the authors.

---

**Fig. A1. Continuous-Time Markov Flow Dynamics In Each Period.**

**Table A1**

Continuous time Markov process analysis — year 1 random assignments.

<table>
<thead>
<tr>
<th>Dependent measure</th>
<th>Continuous time Markov estimation not mediated</th>
<th>Purchase</th>
<th>Continuous time Markov estimation mediated by trust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consider brand</td>
<td></td>
<td>Consider brand</td>
</tr>
<tr>
<td></td>
<td>Not consider to consider (1 → 2)</td>
<td>.139</td>
<td>.146*</td>
</tr>
<tr>
<td></td>
<td>Consider to not consider (2 → 1)</td>
<td>.231</td>
<td>.249*</td>
</tr>
<tr>
<td></td>
<td>Purchase</td>
<td>.120</td>
<td>.102*</td>
</tr>
<tr>
<td></td>
<td>Consider to purchase (2 → 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not consider to consider (1 → 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consider to not consider (2 → 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Purchase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consider to purchase (2 → 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prior ownership of AAM</td>
<td>-.407</td>
<td>-.403</td>
</tr>
<tr>
<td></td>
<td>Prior ownership of American</td>
<td>-.773</td>
<td>-.624</td>
</tr>
<tr>
<td></td>
<td>Prior ownership of Japanese</td>
<td>.070</td>
<td>.565</td>
</tr>
<tr>
<td></td>
<td>Month 3</td>
<td>.012</td>
<td>-.394</td>
</tr>
<tr>
<td></td>
<td>Month 4</td>
<td>-.971</td>
<td>-.394</td>
</tr>
<tr>
<td></td>
<td>Month 5</td>
<td>-.698</td>
<td>-.760</td>
</tr>
<tr>
<td></td>
<td>Log likelihood</td>
<td>-.616.46</td>
<td>-.608.12</td>
</tr>
</tbody>
</table>

All flows are estimated simultaneously.

* Significant at the 0.05 level.

* Significant at the 0.10 level.
The Markov analyses reinforce the conditional-logit and main-effect analyses. Competitive test drives have a significant effect on consideration, but that effect is likely mediated through lagged trust. Lagged trust has a significant effect on key flows. We also modeled potential misclassification of “consider” vs. “not consider” as in Jackson, Sharpes, Thompson, Duffy, and Couto (2003). The misclassification analyses improved fit but provided no additional managerial insights. Estimated misclassification was moderate.

References

Performance implications of deploying marketing analytics

Frank Germann b,1, Gary L. Lilien a,*, Avindr Rangaswamy c,2

a Smeal College of Business, Pennsylvania State University, 484 Business Building, University Park, PA 16802, USA
b University of Notre Dame, 395 Mendoza College of Business, Notre Dame, IN 46556, USA
c Smeal College of Business, Pennsylvania State University, 210E Business Building, University Park, PA 16802, USA

A R T I C L E   I N F O
Article history:
First received in 29, February 2012 and was under review for 3½ months
Available online 27 November 2012
Area Editor: Dominique M. Hanssens

Keywords:
Marketing analytics
Marketing models
Marketing ROI

A B S T R A C T
A few well-documented cases describe how the deployment of marketing analytics produces positive organizational outcomes. However, the deployment of marketing analytics varies widely across firms, and many C-level executives remain skeptical regarding the benefits that they could gain from their marketing analytics efforts. We draw on upper echelons theory and the resource-based view of the firm to develop a conceptual framework that relates the organizational deployment of marketing analytics to firm performance and that also identifies the key antecedents of that deployment. The analysis of a survey of 212 senior executives of Fortune 1000 firms demonstrates that firms attain favorable and apparently sustainable performance outcomes through greater use of marketing analytics. The analysis also reveals important moderators: more intense industry competition and more rapidly changing customer preferences increase the positive impact of the deployment of marketing analytics on firm performance. The results are robust to the choice of performance measures, and, on average, a one-unit increase in the degree of deployment (moving a firm at the median or the 50th percentile of deployment to the 65th percentile) on a 1–7 scale is associated with an 8% increase in return on assets. The analysis also demonstrates that support from the top management team, a supportive analytics culture, appropriate data, information technology support, and analytics skills are all necessary for the effective deployment of marketing analytics.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

A recent Google search for “marketing analytics” returned more than 500,000 hits. Marketing analytics, a “technology-enabled and model-supported approach to harness customer and market data to enhance marketing decision making” (Lilien, 2011, p. 5) consists of two types of applications: those that involve their users in a decision support framework and those that do not (i.e., automated marketing analytics). During the past half century, the marketing literature has documented numerous benefits of the use of marketing analytics, including improved decision consistency (e.g., Natter, Mild, Wagner, & Taudes, 2008), explorations of broader decision options (e.g., Sinha & Zoltners, 2001), and an ability to assess the relative impact of decision variables (e.g., Silk & Urban, 1978). The common theme in this literature is the improvement in the overall decision-making process (e.g., Russo & Schoemaker, 1989, p. 137).

Rapid technological and environmental changes have transformed the structure and content of marketing managers’ jobs. These changes include (1) pervasive, networked, high-powered information technology (IT) infrastructures, (2) exploding volumes of data, (3) more sophisticated customers, (4) an increase in management’s demands for the demonstration of positive returns on marketing investments, and (5) a global, hypercompetitive business environment. In this changing environment, opportunities for the deployment of marketing analytics to increase profitability seemingly should abound. Indeed, an entire stream of research in marketing documents the positive performance implications of deploying marketing analytics (e.g., Hoch & Schkade, 1996; Kannan, Kline Pope, & Jain, 2009; Lodish, Curtis, Ness, & Simpson, 1988; McIntyre, 1982; Natter et al., 2008; Silva-Risso, Bucklin, & Morrison, 1999; Zoltners & Sinha, 2005).

However, there continue to be many skeptics with regard to the “rational analytics approach” to marketing. For example, in a recent interview with one of the authors, a (former) senior executive at one of the world’s leading car manufacturers claimed that “…marketing analytics-based results usually raise more questions than they answer,” and he asserted that “the use of marketing analytics often slows you down.” He also claimed that the “…performance implications of marketing analytics are at best marginal.” When we inquired about documentation for his views, he referred us to Peters and Waterman’s (1982) highly influential book, In Search of Excellence, in which the authors denounce formal analysis because of its abstraction from reality and its tendency to produce “paralysis through analysis” (p. 31). More recently, a study of 587 C-level executives of large international companies revealed that only approximately 10% of the firms regularly employ marketing analytics (McKinsey & Co., 2009). And Kucera and White (2012) note that only 16% of the 160 business
leaders who responded to their survey reported using predictive analytics, although those users “significantly outpace those that do not in two important marketing performance metrics” (p. 1). John Little diagnosed the issue more than 40 years ago as follows: “The big problem with … models is that managers practically never use them. There have been a few applications, of course, but the practice is a pallid picture of the promise” (Little, 1970, p. B-466). Revisiting the issue, Little (2004, p. 1858) reports that “The good news is that more managers than ever are using models … what has not changed is organizational inertia.” Winer (2000, p. 143) concurs: “My contacts in consumer products firms, banks, advertising agencies and other large firms say that [model builders] are a rare find and that models are not used much internally. Personal experience with member firms of MSI indicates the same.”

The low prevalence of marketing analytics use implies that many managers remain unconvinced about the benefits that accrue from that use. In addition, most research studies that document these benefits have focused on isolated firm or business unit “success stories” without systematically exploring performance implications at the firm level. Given the lack of compelling evidence about the performance implications of marketing analytics, the objective of this research is to address two questions: (1) Does widespread deployment of marketing analytics within a firm lead to improved firm performance? and (2) If the answer to (1) is “yes,” what leads to the widespread deployment of marketing analytics within firms? With the usual caveats and cautions, particularly with regard to making causal inferences using non-experimental data, we find that the answer to question 1 appears to be “yes” and, hence, the answer to question 2 has high relevance, as well as academic importance.

To address our research questions, we propose a conceptual framework that relies on both the resource-based view (RBV) of the firm (Barney, 1991; Wernerfelt, 1984) and upper echelons theory (Hambrick & Mason, 1984) to model the factors that link marketing analytics deployment to firm performance, as well as the factors that drive the deployment of marketing analytics. We assess the validity and value of that framework with data drawn from a survey of 212 senior executives at Fortune 1000 firms, supplemented by secondary source objective performance data for those firms. We find that the deployment of marketing analytics has a greater impact on firm performance when the industry is characterized by strong competition and when customer preferences change frequently in the industry. We also find that top management team (TMT) advocacy and a culture that is supportive of marketing analytics are the keys to enabling a firm to benefit from the use of marketing analytics, and our analyses suggest that the benefits realized by marketing analytics deployment may be sustainable.

We proceed as follows: We first present our conceptual framework and hypotheses and, then, describe our data and our methodology. We then present our findings and discuss their theoretical and managerial implications, as well as the limitations of our research.

2. Conceptual framework

The conceptual framework in Fig. 1 depicts what we refer to as the marketing analytics chain of effects. The framework articulates our predicted relationships, including the hypothesized relationship between the deployment of marketing analytics and firm performance.

---

3 The metrics are “incremental lift from a sales campaign” and “click through rate (for mass campaigns).” Those firms that use customer analytics also report a significantly greater ability to measure customer profitability and lifetime value and are also more likely to have staff dedicated to data mining.

4 We use the term “deployment” or “to deploy” to mean “to put into use, utilize or arrange for a deliberate purpose,” without reference to the financial, human, or technical investment that might be necessary for the enablement of such deployment.

We propose that marketing analytics deployment, which we define as the extent to which insights gained from marketing analytics guide and support marketing decision making within the firm, has a positive impact on firm performance. However, this positive impact on firm performance is likely to be moderated by three industry-specific factors: (1) the degree of competition faced by the firm, (2) the rate of change in customer preferences, and (3) the prevalence of marketing analytics use within the industry. Furthermore, we identify TMT advocacy of marketing analytics as a vital antecedent of the deployment of marketing analytics. We suggest that a firm’s TMT must not only commit adequate resources in the form of employee analytic skills, data, and IT but also nurture a culture that supports the use of marketing analytics. Such a culture can ensure that the insights gained from marketing analytics are deployed effectively.

In the following section, we first elaborate on the link between the deployment of marketing analytics and firm performance. Next, we consider the antecedents of the deployment of marketing analytics; i.e., the resources and organizational elements that we posit must be in place for marketing analytics to be deployed effectively.

2.1. The performance implications of deploying marketing analytics

A few authors (primarily authors writing for non-academic journals) suggest that the use of marketing analytics can slow firms down, leading to missed market opportunities that are seized by more agile and non-analytics-oriented competition. For example, citing General Colin Powell’s leadership primer, Harari (1996, p. 37) suggests that “excessive delays in the name of information-gathering breeds analysis paralysis,” which leads to missed opportunities and, hence, subpar firm performance. Peters and Waterman (1982) predict an analogous effect. Additionally, based on our discussions with executives, we conclude that many top managers share similar notions regarding the performance outcomes of marketing analytics use.

However, there are many firm-specific case studies that describe the positive performance impact of marketing analytics use. For example, Elsner, Krafft, and Huchzermeier (2004) demonstrate how Rhenania, a medium-sized German mail order company, used a dynamic, multilevel response modeling system to answer its most important direct marketing questions: When, how often, and to whom should the company mail its catalogs? The model allowed the company to increase its customer base by more than 55% and quadrupled its profitability during the first few years following implementation, and the firm’s president asserted that the firm was saved by deploying this model.

Marketing analytics can also significantly improve a firm’s ability to identify and assess alternative courses of action. For example, in the 1980s, Marriott Corporation was running out of adequate downtown locations for its new full-service hotels. To maintain growth, Marriott’s management planned to locate hotels outside downtown areas to appeal to both business and leisure travelers. A marketing analytics approach called conjoint analysis facilitated the company’s design and launch of its highly successful Courtyard by Marriott chain, establish a multibillion dollar business, and create a new product category (Wind, Green, Shifflet, & Scarbrough, 1989).

In another example, Kannan, Kline Pope, and Jain (2009) report how marketing analytics at the National Academies Press (NAP) led to a better understanding of customers and to a better manner of reaching the customers. The NAP was concerned about the best way to price and distribute its book in print and in pdf format via the Internet. It built a pricing model that allowed for both substitution and complementarity effects among the two formats and calibrated the model using a choice modeling experiment. The results permitted the NAP to launch its entire range of digital products with a variable pricing scheme, thereby maximizing the reach of its authors’ work.

The common theme of the above firm-specific examples is that the deployment of marketing analytics allows firms to develop and offer
products and services that are better aligned with customer needs and wants, which, in turn, leads to improved firm performance. Thus, we propose the following main effect:

**H1.** The greater the deployment of marketing analytics, the better the firm’s performance.

### 2.1.1. Competitive industry structure

Most firms compete with a number of rivals (Debruyne & Reibstein, 2005), although the degree of rivalry varies considerably across industries (DeSarbo, Grewal, & Wind, 2006). The level of competition that a firm faces also has many concomitant effects, including the degree of customer satisfaction that the firm must attain to operate successfully. For example, Anderson and Sullivan (1993) find that firms with less satisfied customers that face less competition perform approximately the same or even better than firms with more satisfied customers that operate in more competitive environments. Thus, firms that confront more competition must strive for higher levels of customer satisfaction to perform well.

Assuming that marketing analytics provide better insights about customer needs, firms in industries with greater competition should earn higher returns (because of more clearly targeted offerings, for example, which result in greater customer satisfaction) than firms in less competitive industries. Thus, we propose:

**H2.** The more intense the level of competition among industry participants, the greater the positive impact of marketing analytics deployment on firm performance.

We note that if “analysis–paralysis” is a serious concern associated with the deployment of marketing analytics, then the corresponding negative performance implications should be even greater in competitive environments because competitors move more swiftly in such environments (e.g., DeSarbo et al., 1996). Under these circumstances, we should observe a negative interaction between marketing analytics deployment and level of competition (as opposed to our predicted positive interaction).

### 2.1.2. Customer preference changes

Customer preferences regarding product features, price points, distribution channels, media outlets, and other elements of the marketing mix change over time (e.g., Kotler & Keller, 2006, p. 34). The rate of such change varies: fashions change seasonally, whereas preferences for consumer electronics appear to change almost monthly (e.g., Lamb, Hair, & McDaniel, 2009, p. 58), but preferences regarding construction equipment, hand tools, and agricultural products appear to be much more stable over time.

The more customers’ needs fluctuate, the greater is the uncertainty that firms face in making decisions and the more critical scanning and interpreting the changing environment becomes (Daft & Weick, 1984). Marketing analytics offer various means to assist firms in monitoring the pulse of the market and providing early warning of preference changes. Additionally, a stable, predictable environment reduces the need for marketing analytics because such an environment requires a limited number of decision variables to manage for organizational success (Smart & Vertinsky, 1984). Therefore, we propose:

**H3.** The more rapidly customer preferences change in an industry, the greater the positive impact of the deployment of marketing analytics on firm performance.

### 2.1.3. Prevalence of marketing analytics use

The prevalence of the use of marketing analytics within an industry may attenuate their positive performance implications. Porter (1996, p. 63) notes that as firms evolve, “staying ahead of rivals gets harder,” partially because of the diffusion of best practices, facilitated, for example, by inputs from strategy consultants. Competitors are quick to imitate successful management techniques, particularly if they promise superior methods of understanding and meeting customers’ needs. Such imitation eventually raises the bar for everyone (e.g., Chen, Su, & Tsai, 2007; D’Aveni, 1994; MacMillan, McCaffery, & Van Wijk, 1985). Thus, the greater the overall use of marketing analytics in an industry, the lower is the upside potential for a firm to increase its use. Hence, we propose:

**H4.** The more prevalent the use of marketing analytics in an industry, the lower is the positive impact of the deployment of marketing analytics on the performance of individual firms in that industry.

To summarize our hypotheses regarding research question #1, we predict that the deployment of marketing analytics has positive performance implications in general and that this effect is even stronger...
in industries characterized by strong competition and in which customer preferences change frequently and weaker in industries in which the deployment of marketing analytics is commonplace.

We next discuss the factors that lead to the deployment of marketing analytics.

2.2. Antecedents of the deployment of marketing analytics

Adapting a resource-based view (RBV—Barney, 1991; Wernerfelt, 1984), Amit and Schoemaker (1993) suggest that firms create competitive advantage by assembling, integrating, and deploying their resources in a manner that allows them to work together to create firm capabilities. Firm capabilities can provide a sustainable competitive advantage when they are protected by isolating mechanisms that thwart competitive imitation (Rumelt, 1984).

Building on the RBV literature, we suggest that marketing analytics must be appropriately assembled and embedded within the fabric of the firm to be deployed effectively, which potentially results in a sustainable competitive advantage. Furthermore, we single out TMT advocacy of marketing analytics as a key driver of that process.

2.2.1. TMT advocacy, analytics culture, and sustainable competitive advantage

According to upper echelons theory (Hambrick & Mason, 1984), organizations are a reflection of their TMT; thus, for marketing analytics to become an integral part of a firm’s business routines and, ultimately, its culture, it must be strongly supported by the firm’s TMT (Hambrick, 2005).

We posit that a culture that is supportive of marketing analytics is critical for its effective deployment because that culture carries the logic of how and why “things happen” (Deshpande & Webster, 1989, p. 4). These norms are especially important because the person (or organizational unit) that carries out the marketing analytics (e.g., marketing analyst or researcher) frequently is not responsible for implementing the insights gained, namely, executives in marketing and other functions (Carlsson & Turban, 2002; Hoekstra & Verhoef, 2011; Van Bruggen & Wierenga, 2010; Wierenga & van Bruggen, 1997). An analytics culture provides decision makers with a pattern of shared values and beliefs (Deshpande, Farley, & Webster, 1993; Ouchi, 1981), which in turn, should positively influence the degree to which they incorporate the insights gained from marketing analytics in their decisions. Furthermore, culture is sticky, difficult to create, and even more difficult to change (e.g., Schein, 2004), suggesting that it may protect against competitive imitation of a firm’s analytics investments, thus delivering sustainable rewards from a firm’s marketing analytics investments.

2.2.2. Analytics skills

To deploy marketing analytics within a firm, the firm must also have access to people (either internally or among its partners) who have the knowledge to execute marketing analytics. Thus, the TMT must ensure that people with the requisite marketing analytics skills are present within the company or available outside the firm. We distinguish between technical marketing analytics skills and other individual-level, analytics-based knowledge structures that are tacit (Grant, 1991). Technical marketing analytics skills likely derive primarily from classroom or other structured learning situations and consist of the range of marketing models and related concepts that the analyst could deploy. In contrast, tacit knowledge of marketing analytics includes skills acquired primarily through real-world learning.

We anticipate that higher levels of marketing analytics skills will increase the extent of marketing analytics deployment because people use the tools and skills they understand and with which they are comfortable (Lounsbury, 2001; Westphal, Gulati, & Shortell, 1997). Additionally, better skills should lead to more useful results from using those skills, thus facilitating the organization-wide marketing analytics adoption process. Therefore, a firm’s employees’ analytics skills should have both a direct, positive impact on the organizational deployment of analytics and an indirect effect on organizational deployment through the positive impact on analytics culture.

2.2.3. Data and IT resources

A firm’s physical IT infrastructure and data resources are two other critical tangible assets that the TMT must implement to allow for the effective deployment of marketing analytics. Physical IT resources form the core of a firm’s overall IT infrastructure and include computer and communication technologies and shared technical platforms and databases (Ross, Beath, & Goodhue, 1996). Data result from measurements and provide the basis for deriving information and insights from marketing analytics (Lilien & Rangaswamy, 2008). Marketing analytics are often based on vast amounts of customer data (Roberts, Morrison, & Nelson, 2004), which require sophisticated IT resources to effectively obtain, store, manipulate, analyze, and distribute across the firm. Therefore, IT and data are closely related tangible resources, such that one would be significantly less valuable without the other. Building on this mutual dependence, we posit that both IT and data resources are important prerequisites for marketing analytics use.

To summarize our hypotheses regarding research question #2, we propose that TMT advocacy of marketing analytics is an important precursor to the effective deployment of marketing analytics. We further propose that a firm’s TMT must not only ensure that employees with the requisite analytics skills and an adequate data and IT infrastructure are in place but also nurture a culture that supports the use of marketing analytics. Such a culture can ensure that the insights gained from marketing analytics are deployed effectively.

3. Data and methods

3.1. Scale development

We adapted existing scales when they were available. However, our study is among the first to empirically explore the performance implications of marketing analytics, and scales for several of our constructs were not available. We developed the missing scales, following a four-phase iterative procedure, as recommended in the literature (Churchill, 1979): First, we independently generated a large pool of items for each of the constructs from an extensive literature review. Second, we engaged fifteen senior-level, highly regarded marketing academics to expand our list of items and evaluate the clarity and appropriateness of each item. Third, we personally administered pretests to six top managers to assess any ambiguity or difficulty that they experienced when responding. Fourth, we conducted a formal pretest with 31 senior managers. Because the fourth stage/pretest revealed no additional concerns, we finalized the scale items, which are listed in Appendix A.6

3.2. Data collection procedure

We conducted a mail survey among executives of Fortune 1000 firms. We first randomly selected 500 entries from the Fortune 1000 firms characterized by strong competition and in which customer preferences change frequently and weaker in industries in which the deployment of marketing analytics is commonplace.

We next discuss the factors that lead to the deployment of marketing analytics.

Adapting a resource-based view (RBV—Barney, 1991; Wernerfelt, 1984), Amit and Schoemaker (1993) suggest that firms create competitive advantage by assembling, integrating, and deploying their resources in a manner that allows them to work together to create firm capabilities. Firm capabilities can provide a sustainable competitive advantage when they are protected by isolating mechanisms that thwart competitive imitation (Rumelt, 1984).

Building on the RBV literature, we suggest that marketing analytics must be appropriately assembled and embedded within the fabric of the firm to be deployed effectively, which potentially results in a sustainable competitive advantage. Furthermore, we single out TMT advocacy of marketing analytics as a key driver of that process.

2.2.1. TMT advocacy, analytics culture, and sustainable competitive advantage

According to upper echelons theory (Hambrick & Mason, 1984), organizations are a reflection of their TMT; thus, for marketing analytics to become an integral part of a firm’s business routines and, ultimately, its culture, it must be strongly supported by the firm’s TMT (Hambrick, 2005).

We posit that a culture that is supportive of marketing analytics is critical for its effective deployment because that culture carries the logic of how and why “things happen” (Deshpande & Webster, 1989, p. 4). These norms are especially important because the person (or organizational unit) that carries out the marketing analytics (e.g., marketing analyst or researcher) frequently is not responsible for implementing the insights gained, namely, executives in marketing and other functions (Carlsson & Turban, 2002; Hoekstra & Verhoef, 2011; Van Bruggen & Wierenga, 2010; Wierenga & van Bruggen, 1997). An analytics culture provides decision makers with a pattern of shared values and beliefs (Deshpande, Farley, & Webster, 1993; Ouchi, 1981), which in turn, should positively influence the degree to which they incorporate the insights gained from marketing analytics in their decisions. Furthermore, culture is sticky, difficult to create, and even more difficult to change (e.g., Schein, 2004), suggesting that it may protect against competitive imitation of a firm’s analytics investments, thus delivering sustainable rewards from a firm’s marketing analytics investments.

2.2.2. Analytics skills

To deploy marketing analytics within a firm, the firm must also have access to people (either internally or among its partners) who have the knowledge to execute marketing analytics. Thus, the TMT must ensure that people with the requisite marketing analytics skills are present within the company or available outside the firm. We distinguish between technical marketing analytics skills and other individual-level, analytics-based knowledge structures that are tacit (Grant, 1991). Technical marketing analytics skills likely derive primarily from classroom or other structured learning situations and consist of the range of marketing models and related concepts that the analyst could deploy. In contrast, tacit knowledge of marketing analytics includes skills acquired primarily through real-world learning.

We anticipate that higher levels of marketing analytics skills will increase the extent of marketing analytics deployment because people use the tools and skills they understand and with which they are comfortable (Lounsbury, 2001; Westphal, Gulati, & Shortell, 1997). Additionally, better skills should lead to more useful results from using those skills, thus facilitating the organization-wide marketing analytics adoption process. Therefore, a firm’s employees’ analytics skills should have both a direct, positive impact on the organizational deployment of analytics and an indirect effect on organizational deployment through the positive impact on analytics culture.

2.2.3. Data and IT resources

A firm’s physical IT infrastructure and data resources are two other critical tangible assets that the TMT must implement to allow for the effective deployment of marketing analytics. Physical IT resources form the core of a firm’s overall IT infrastructure and include computer and communication technologies and shared technical platforms and databases (Ross, Beath, & Goodhue, 1996). Data result from measurements and provide the basis for deriving information and insights from marketing analytics (Lilien & Rangaswamy, 2008). Marketing analytics are often based on vast amounts of customer data (Roberts, Morrison, & Nelson, 2004), which require sophisticated IT resources to effectively obtain, store, manipulate, analyze, and distribute across the firm. Therefore, IT and data are closely related tangible resources, such that one would be significantly less valuable without the other. Building on this mutual dependence, we posit that both IT and data resources are important prerequisites for marketing analytics use.

To summarize our hypotheses regarding research question #2, we propose that TMT advocacy of marketing analytics is an important precursor to the effective deployment of marketing analytics. We further propose that a firm’s TMT must not only ensure that employees with the requisite analytics skills and an adequate data and IT infrastructure are in place but also nurture a culture that supports the use of marketing analytics. Such a culture can ensure that the insights gained from marketing analytics are deployed effectively.

3. Data and methods

3.1. Scale development

We adapted existing scales when they were available. However, our study is among the first to empirically explore the performance implications of marketing analytics, and scales for several of our constructs were not available. We developed the missing scales, following a four-phase iterative procedure, as recommended in the literature (Churchill, 1979): First, we independently generated a large pool of items for each of the constructs from an extensive literature review. Second, we engaged fifteen senior-level, highly regarded marketing academics to expand our list of items and evaluate the clarity and appropriateness of each item. Third, we personally administered pretests to six top managers to assess any ambiguity or difficulty that they experienced when responding. Fourth, we conducted a formal pretest with 31 senior managers. Because the fourth stage/pretest revealed no additional concerns, we finalized the scale items, which are listed in Appendix A.6

3.2. Data collection procedure

We conducted a mail survey among executives of Fortune 1000 firms. We first randomly selected 500 entries from the Fortune 1000
list and then leveraged the corporate connections of two major U.S. universities to obtain the names of 968 senior executives (primarily alumni) working at these firms.

We addressed these respondents using personalized letters, in which we asked them to complete the survey in reference to either their strategic business unit (SBU) or their company, whichever they felt was more appropriate. We also provided a nominal incentive (1 USD, called a token of thanks, which emerged as the most effective incentive in a pretest). Of the 968 executives contacted, 36 returned the surveys and indicated they were not qualified to respond and 20 surveys were returned because of incorrect addresses. We obtained 212 completed surveys (of the 912 remaining surveys), which yielded an effective response rate of 23.25%. We controlled for possible nonresponse bias by comparing the construct means for early and late respondents (Armstrong & Overton, 1977) but found no significant differences. As we show in Table 1, most (71%) of the respondents in our sample had titles of director or higher, which suggests that they should be knowledgeable about their firms’ capabilities and actions.

We also asked the respondents to report their confidence levels with regard to the information they provided (Kumar, Stern, & Anderson, 1993). The sample mean score was 5.59 (out of 7 [SD = .81]), indicating a high level of confidence. Additionally, we received multiple (either two or three) responses from 35 firms/SBUs in our sample, allowing us to cross-check the responses when we received more than one response from a firm.1

### 3.3. Scale assessment

We assessed the reliability and validity of our constructs using confirmatory factor analysis (Bagozzi, Yi, & Phillips, 1991; Gerbing & Anderson, 1988). We included all independent and dependent latent variables in one confirmatory factor analysis model, which provided satisfactory fit to the data (comparative fit index [CFI] = .97; root mean square error of approximation [RMSEA] = .05; 90% confidence interval [CI] of RMSEA = [.033; .068]). On the basis of the estimates from this model, we examined the composite reliability and discriminant validity of our constructs (Fornell & Larcker, 1981). All composite reliabilities exceed the recommended threshold value of .6 (Bagozzi & Yi, 1988); the lowest reliability is .75. The coefficient alphas of our constructs are all greater than .7. We also assessed discriminant validity using the criteria proposed by Fornell and Larcker (1981). The results demonstrate that the squared correlation between any two constructs is always lower than the average variance extracted (AVE) for the respective constructs, providing support for discriminant validity. Finally, the correlations between the respective constructs are all significantly different from unity (Gerbing & Anderson, 1988). Overall, the results indicate that our latent constructs demonstrate satisfactory levels of composite reliability and discriminant validity. We present the correlations among the constructs in Table 2 and the AVE and coefficient alphas in the Appendix A along with the scale items.

Although we were able to establish discriminant validity, some of our constructs are highly correlated. For example, the correlation between analytics skills and analytics culture is 0.825. As per our measures, “analytics skills” refer to the type of analytics skills that the employees possess, whereas “analytics culture” indicates shared beliefs with regard to how analytics will influence the company. Although one would expect these two constructs to be highly correlated, we assert that they do not measure the same thing, much in the same manner that a physician who measures a patient’s height and weight, two highly correlated items, might argue that height and weight measure different important things and thus both should be measured.

#### 3.3.1. Descriptive statistics

Table 3 contains descriptive statistics for our sample firms and indicates that the sample represents a broad range of firms. Table 4 lists the names of some sample firms. In Table 5, we provide the summary statistics and correlations for our variables and, in Table 6, we present histograms for our focal variables. As the histograms show, the sampled firms display a wide range of values for our focal variables. For example, on the seven-point scale measuring TMT advocacy of marketing analytics, approximately 18% of the sample firms fall within the 6–7 range and 16% within the 1–3 range (M = 4.5; SD = 1.7). Furthermore, with regard to analytics culture, approximately 25% of the sample firms fall within the 6–7 range, and approximately 14% score within the 1–3 range (M = 4.6; SD = 1.6). We also asked the respondents (1) whether their marketing analytics applications are designed primarily in-house or by outside experts and (2) whether the primary day-to-day operations of marketing analytics are managed in-house or outsourced. Table 7 presents the responses to these questions and demonstrates that the majority of the Fortune 1000 firms design and manage their marketing analytics (applications) in-house. We also make note of the low percentage of respondents who did not know the answer to these questions, another sign that our respondents are quite knowledgeable about the domain under study.

### 3.4. Conceptual model testing procedures

Our conceptual model proposes both direct and moderating effects (Fig. 1). To model and test these effects simultaneously, we used structural equation modeling (SEM); recent methodological advances have made it feasible to include multiple interactions in a path model (Klein & Moosbrugger, 2000; Marsh, Wen, & Hau, 2004; Muthén & Asparouhov, 2003). We used Mplus Version 6.11 and estimated our model using the full-information maximum likelihood approach (Klein & Moosbrugger, 2000; Muthén & Muthén, 2010, p. 71).

### 4. Results

#### 4.1. SEM model fit

Fig. 2 summarizes the results of our SEM, depicting two of the three interactions (i.e., competition and needs and wants change) as statistically significant. Because means, variances, and covariances are not sufficient statistics for our SEM estimation approach, our model does not provide the commonly used fit statistics, such as RMSEA and CFI. Instead, in accordance with Muthén (2010), we assessed fit in two steps. First, we re-estimated our SEM without the interaction terms and compared that model with our original model via a chi-square difference test using the associated loglikelihoods (Muthén & Muthén, 2011; Satorra & Bentler, 1999). This test yielded a $\chi^2 (3)$ difference of

### Table 1

<table>
<thead>
<tr>
<th>Position</th>
<th>Number of participants</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>President, CEO</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>EVP, (Sr.) VP, CMO, CFO, COO</td>
<td>78</td>
<td>37</td>
</tr>
<tr>
<td>(Sr.) Director, Executive Director</td>
<td>65</td>
<td>31</td>
</tr>
<tr>
<td>(Sr.) Marketing Manager</td>
<td>47</td>
<td>22</td>
</tr>
<tr>
<td>Other (e.g., Marketing Strategist)</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>212</td>
<td>100</td>
</tr>
</tbody>
</table>

1. We received two responses from 33 firms/SBUs and three responses from 2 firms/SBUs. Because we had contacted 968 executives who worked for 500 randomly selected Fortune 1000 firms, we evidently contacted multiple executives working for the same firms/SBUs, which accounts for most of these multiple responses. In a few instances ($n = 5$), executives also invited their coworkers to participate in the survey.

2. Although this multiple-response sample is too small for a formal multivariate multmethod assessment, it enabled us to assess whether the respective respondent groups’ means for the key constructs were statistically different (e.g., Srinivasan, Lilien, & Rangaswamy, 2002). T-tests indicated that none of the means were statistically significantly different from each other.
28.124, which is highly significant ($p < .0001$) which clearly favors the model with interactions. Second, we (re)estimated the model without interactions with the conventional SEM estimation approach to derive the usual model fit statistics (e.g., RMSEA and CFI). This conventional model (without interactions) fits the data quite well ($\chi^2 (175) = 243$; CFI = .97; RMSEA = .04; 90% C.I. = [.03; .06]), and the paths are very similar to those of the moderated model. Based on these results, we conclude that the “un-moderated” model fits the data well and that the moderated model enhances the model fit.

4.2. Specific model paths and hypothesis test results

All of the paths from TMT advocacy to the respective subsequent latent constructs are positive and significant, suggesting that the TMT plays a key role in establishing an organizational setting in which marketing analytics can be deployed effectively. Additionally, as predicted, an analytics-oriented culture has a positive and significant effect on the deployment of analytics ($\beta = .317$, $p < .01$), in line with our proposition that strengthening a firm’s analytics-oriented culture leads to an actual increase in the deployment of marketing analytics. In addition, we find that enhancements to a firm’s marketing analytics skills have both a direct and positive impact on the deployment of analytics ($\beta = .427$, $p < .001$) and a positive, indirect effect through analytics culture ($\beta = .120$, $p < .05$). That is, employees’ marketing analytics skills directly influence the degree to which the firm uses analytics-based findings in marketing decision making; they also exert an indirect influence by enhancing the organization’s analytics-oriented culture. We also find that the presence of a strong data and IT infrastructure promotes marketing analytics skills within the firm ($\beta = .621$, $p < .001$).

As hypothesized in H$_4$, higher levels of deployment of marketing analytics leads to an increase in firm performance ($\beta = .106$, $p < .01$). Moreover, as hypothesized in H$_5$, we find both a positive and significant deployment of analytics $\times$ competition interaction ($\beta = .081$, $p < .05$), which shows that the use of analytics is more effective in more competitive environments than in less competitive environments.$^{10}$ Similarly, in support of H$_6$, the use of analytics is more effective in environments in which customers’ needs and wants change frequently ($\beta = .060$, $p < .01$). However, we do not find support for H$_7$ concerning the analytics $\times$ prevalence interaction ($\beta = -.034$, ns).

4.3. Robustness checks

4.3.1. Validity of the performance measure/monomethod bias

Because our independent and dependent measures originate from the same respondents, leading to the possibility of monomethod bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), we collected performance data from independent sources to validate our performance measure. We obtained information on firm-specific net income and total assets for as many firms as possible by retrieving their 10-K and other filings with the U.S. Securities and Exchange Commission

---

Note: The correlations and their standard errors (provided in brackets underneath) are in bold, the squared correlations are in italics, and the variances are provided on the diagonal.
from the EDGAR database. We also consulted COMPUSTAT, Merger Online and the firms’ websites. With these financial data, we computed the respective firm’s return on assets (ROA). These procedures yielded financial performance data for 68 of the 212 responses. After matching the time horizon of the performance measures, we computed a 2-year average ROA for the 2 years preceding our primary data collection (see, for example, Boulding, Lee, & Staelin, 1994). We also standardized the ROA measure with respect to each firm’s competitors (from Merger Online).

To address same-source bias, we used the objective performance data (i.e., ROA) to reanalyze our conceptual framework. Given the small sample size and the consequent lack of statistical power (𝑛 = 68), it was not feasible to simultaneously test all of the hypothesized effects of our framework in a single SEM model. Instead, we conducted two separate analyses: first, we used a SEM to estimate the direct (un-moderated) effects in our conceptual framework. Second, we used an ordinary least squares (OLS) regression model to retest Hu–H6. We substituted the ROA objective performance measure for the perceptual performance measure in both analyses.

The SEM results remain consistent regardless of the use of objective or subjective data; in fact, the link from deployment to performance is even stronger with objective data than with subjective data. We provide the SEM results with objective data in Fig. 3.

We report the regression results with objective data in Table 8 (model 1). We used a simple average of the items measuring deployment of analytics as our deployment construct in that analysis. We repeated the analyses using the factor scores from our SEM for our deployment construct. These two measures were highly correlated (correlation > .94), and none of our inferences were affected by the choice of deployment construct. Overall, the regression model is significant, and our inferences did not change.

In summary, the signs of the SEM and regression model coefficients using objective data are consistent with those obtained using the survey-based data. However, the deployment of analytics × competition interaction did not reach significance in the regression model (𝑡 = 1.60), a result that could be due to the small sample size for the objective data (𝑛 = 68).

4.3.2. Multiple respondents for some firms

As noted, we obtained data from multiple respondents from 35 organizational units. To address potential issues of non-independence among these observations in our data, we averaged the responses of multiple respondents11 from each firm (e.g., Homburg, Grodzanovic, & Klamann, 2007) and then re-estimated the SEM using individual responses as if we had only obtained single responses (i.e., the average responses for those organizational units for which we obtained multiple responses). The results remain virtually the same, and our inferences do not change.

4.3.3. Multigroup analysis — B2B vs. B2C

There are many differences between business-to-business (B2B) and business-to-consumer (B2C) firms (see Grewal & Lilien, 2012) that might lead one to expect that there would be differences in the role and impact of marketing analytics within B2B and B2C firms. To test this possibility, we performed a multigroup confirmatory factor analysis to compare the factor loadings of B2B with B2C firms. To test for partial measurement invariance across groups, we compared a model in which all parameters could be unequal across the two groups with one in which we constrained the factor loadings to be equal. The model with all parameters freely estimated fit the data well (χ² (252) = 321.541; CFI = .97; RMSEA = .05), as did the partial invariance model with factor loadings constrained to be equal (χ² (270) = 336.227; CFI = .97; RMSEA = .05). Furthermore, the χ² difference test indicated that the two models were not statistically significantly different (χ² (18) = 14.7, 𝑝 = .68), thereby suggesting that our findings hold across different types of firms.

11 The 𝑡-tests of the key variables across these respondents’ reports indicated that the respective means were not statistically different.
vey. We followed the same procedure as outlined earlier to collect the additional objective performance data for the year following our survey. To (at least partially) assess this potential reverse-causality issue, we collected the deployment of marketing analytics, and not vice versa. To (at least partially) assess this potential reverse-causality issue, we collected the deployment of marketing analytics, and not vice versa. To (at least partially) assess this potential reverse-causality issue, we collected the deployment of marketing analytics, and not vice versa.

4.3.4. Robustness of the deployment to performance link

Our study reveals a statistically significant positive relationship between the deployment of marketing analytics and firm performance (both subjective and objective). This result is of great managerial importance, and, therefore, we subjected this relationship to additional scrutiny via (1) testing for the linearity of this relationship, (2) assessing the effects of various controls, (3) subjecting it to a reverse-causality test, (4) assessing the contemporary vs. carryover effects of deployment on performance, (5) testing for the effects of unobserved heterogeneity, and (6) assessing the unidimensionality of our performance construct. We elaborate on these robustness tests below.

First, we ran an OLS regression model similar to that reported in Table 8 and included a quadratic term to check for curvilinear effects of the deployment of analytics. The squared term was not statistically significant, suggesting the absence of a curvilinear effect, at least within the range of our data.

Second, we included organization size (number of employees) and industry dummy variables as controls in the regression model. Firm size would account for the fact that larger firms could benefit from economies of scale and scope, rendering their use of analytics more effective. Industry dummies would account for differences in industry segments. We used standard industrial classifications to group the sample firms into five categories (see Table 3): services, manufacturing, finance/insurance, trade, and construction/mining. The size and industry dummy variables had neither a main nor a moderating effect on the relationship between deployment of analytics and firm performance, and our inferences did not change. Thus, our results appear robust to firm size and industry segments.

Third, it might be that firms that perform well have more leeway and, hence, more resources to deploy marketing analytics than do those that perform poorly, implying that firm performance may affect the deployment of marketing analytics, and not vice versa. To (at least partially) assess this potential reverse-causality issue, we collected additional objective performance data for the year following our survey. We followed the same procedure as outlined earlier to collect the objective performance data and then calculated the 2-year average ROA using the newly collected data, as well as the data for the year preceding our primary data collection. We then used this new objective performance data to reanalyze our conceptual model. As before, we relied on SEM to estimate the direct (un-moderated) effects in our conceptual model and used OLS regression to examine the link between deployment of marketing analytics and firm performance. We report the SEM results in Fig. 4 and include the regression results in Table 8 (model 2). As the results show, the outcomes did not change in any substantive manner, providing support for the notion that the deployment of marketing analytics is an antecedent of firm performance, not vice versa.

Fourth, to assess the timing of the performance effects of deployment of marketing analytics, we combined the objective performance measures as follows:

\[(\lambda \times \text{Performance}_{\text{time 1}}) + ([1-\lambda] \times \text{Performance}_{\text{time 2}})\]

where \(\lambda\) can range from 0 to 1, \(\text{Performance}_{\text{time 1}}\) is our initial objective ROA measure and \(\text{Performance}_{\text{time 2}}\) is the ROA measure with a 1-year lag. We then re-estimated our OLS regression model, with the resulting linear combination values as the dependent variable (with \(\lambda\) varying in increments of 0.1 from 0 to 1), and assessed which linear combination yields the best fitting model as determined by Adj. \(R^2\). Fig. 5 provides the results of our analyses.

The results reveal that the highest Adj. \(R^2\) occurs when \(\lambda = .4\) (this is the maximum likelihood estimate for \(\lambda\) assuming Normal distribution of the error terms of the OLS regression), suggesting that the performance effects of the deployment of analytics appear to be observed both immediately and with a slightly stronger carryover. This finding further discounts the possibility of a reverse-causality effect, with the effects being slightly stronger in Time 2 than in Time 1 (A value of \(\lambda = .5\) would indicate that the short-term and longer-term effects are the same).

Fifth, we estimated a mixture regression model (DeSarbo & Cron, 1988) to explore the possibility of unobserved heterogeneity among firms. The lowest Bayesian Information Criterion (BIC) emerged for...
a one-class model (consistent with our multi-group analysis above), which suggests that unobserved heterogeneity was not relevant for our model. Thus, our findings appear to be generalizable to all types of Fortune 1000 firms.

Sixth, the correlations among the subjective performance measures (items 16–18 in Table 5) suggest that our performance construct may not be unidimensional: the correlation between profits and return on investment (ROI) is quite high ($r = .832$), whereas the correlations between sales growth and profits ($r = .451$) and sales growth and ROI ($r = .496$) are significantly lower. Therefore, we analyzed the effect of the deployment of analytics on performance with regard to sales growth and profits/ROI separately. In the SEM model, the main effect of the deployment of analytics on performance increased in both instances, i.e., when using only the single-item sales growth measure ($\beta = .171$ vs. .106) and when using the construct comprised of the profits and ROI items ($\beta = .198$ vs. .106). Furthermore, when employing sales growth as the outcome measure, competition no longer emerges as a significant moderator of analytics deployments’ effect on performance ($\beta_{\text{Deployment} \times \text{competition}} = .063$ vs. .081; the interaction between needs and wants change and deployment of analytics remains marginally significant: $\beta_{\text{Deployment} \times \text{needs and wants change}} = .076$ vs. .06). In contrast, both interactions, i.e., competition $\times$ deployment of analytics and needs and wants change $\times$ deployment of analytics become stronger when including the profits/ROI performance variables in the SEM ($\beta_{\text{Deployment} \times \text{competition}} = .149$ vs. .081 and $\beta_{\text{Deployment} \times \text{needs and wants change}} = .113$ vs. .06). All other paths remain virtually the same in the respective models.

Thus, although the use of marketing analytics appears to positively affect sales growth, profits and ROI, our analysis suggests that the deployment of analytics may have a somewhat stronger effect on profits/ROI than on sales growth. We offer the following possible explanations for this finding: First, many marketing analytics applications are geared toward identifying the most profitable customer segment(s) (e.g., Reinzart & Kumar, 2000), applications designed to improve profits and ROI, as opposed to sales. Second, our sample is drawn from Fortune 1000 firms — all large firms — and their scale may prevent them from growing as quickly as smaller firms. Thus, this finding may be specific to our sample and should be explored more broadly.

Table 9 summarizes our robustness checks of the deployment to performance link.

4.3.5. Deployment of analytics as mediator

Our conceptual model assumes that the deployment of analytics mediates the effect of analytics culture and analytics skills on firm performance. To test this assumption, we conducted a formal test of mediation, following the procedure recommended by Baron and Kenny (1986). We used both of the objective performance measures as the respective dependent variables, deployment of analytics as the mediator, and analytics skills or analytics culture as the respective independent variables. Deployment of analytics emerges as a mediator for both independent variables irrespective of the objective performance measure used.

5. Discussion and conclusions

Our research objective was to determine whether the deployment of marketing analytics leads to improved firm performance and to identify the factors that lead firms to deploy marketing analytics. Our findings address these two research objectives and provide insights of value for both marketing theory and practice.

5.1. Theoretical implications

Our study helps explain what drives the adoption of marketing analytics by firms and why that adoption leads to improved firm performance.

We find support for our hypotheses that the positive effect of marketing analytics deployment on firm performance is moderated by the level of competition that a firm faces, as well as by the degree to which the needs and wants of its customers change over time. However, contrary to our hypothesis, the prevalence of marketing analytics use in a given industry does not moderate the effect of marketing analytics on firm performance. We suggest a possible explanation for this (non)result: consistent with McKinsey & Co.’s (2009) findings, the prevalence of marketing analytics use in the industries that we examined is relatively low. That is, the average response of executives who participated in our survey to the statement “marketing analytics are used extensively in our industry” was a 3.4 on a 7-point scale (SD = 1.6). Perhaps the moderating effect of marketing analytics’ prevalence does not emerge until the industry-wide use of marketing analytics reaches a higher level than evidenced in our sample. Our data simply may not provide the necessary range to manifest such an effect, or an issue we plan to examine in more detail in the future. An alternative explanation for the non-significant interaction could be that competitors cannot compete away a firm’s marketing analytics capability that is implemented properly.

We posit and show empirically that a firm’s TMT must ensure that the firm (1) employs people with requisite analytics skills, (2) deploys sophisticated IT infrastructure and data, and (3) develops a culture that supports marketing analytics so that the insights gained from marketing analytics can be deployed effectively within the firm. The people who perform marketing analytics (e.g., marketing analysts) are frequently not those who implement the insights gained from marketing analytics (e.g., marketing executives), but both groups should support the use of marketing analytics if the firm is to possess a strong marketing analytics-oriented culture (Deshpande et al., 1993). Therefore, a suitable analytics culture that promotes the use of marketing analytics is a critical component of our framework. Additionally, the centrality of an analytics culture, which is sticky and difficult to change or replicate, suggests that the deployment of marketing analytics may provide the necessary firm capability properties that can lead to a sustainable competitive advantage (Barney, 1991).

5.2. Managerial implications

Our findings offer several useful implications for managerial practice. First, the low prevalence of marketing analytics use indicates that few managers are convinced of the benefits of marketing analytics. However, our results suggest that most firms can expect favorable performance outcomes from deploying marketing analytics. Moreover, these favorable performance outcomes should be even greater in industries in which competition is high and in which customers change their needs and wants frequently.

The use of objective performance data as the dependent variable in our regression model enables us to quantify the actual performance implications of, for instance, a one-unit increase (on a scale of 1 to 7) in marketing analytics deployment. Consider Firm A in our sample, which is at the median (50th percentile) in deployment of marketing analytics and operates in an industry characterized by average competition and average changes in customer needs and wants. For Firm A, a one-unit increase in the deployment of marketing analytics is associated with an 8% increase in ROA. Now, consider Firm B in our sample, which is also at the median (50th percentile) deployment of marketing analytics but which operates in highly competitive industries with frequently changing customer needs and wants. For Firm B, a one-unit increase is associated with a 21% average increase

---

12 We also examined potential curvilinear effects of marketing analytics prevalence but did not find any such effects.
Table 6
Histograms of focal variables.

Note: (Combined) signifies that the graph reports the average scores of the variables that form the respective latent variables. As the histograms illustrate, the firms in the sample display a wide range of values for our focal variables.

Table 7
Locus of marketing analytics development and execution.

Locus of Marketing Analytics Development and Execution

“Are your marketing analytics applications designed primarily in-house, or by outside experts/consultants?”

“Are the primary DAY-TO-DAY OPERATIONS of the marketing analytics managed in-house, or are they outsourced?”

1 = Primarily in-house; 2 = Primarily external; 3 = Combination of in-house and external; 4 = Don’t know.
in ROA. The 8% increase in ROA translates to an expected increase of approximately $70 million in net income for the firms in our sample; the 21% increase indicates an increase of $180 million in net income.14

Second, if implemented properly, the use of marketing analytics could be a source of a sustainable competitive advantage for a firm. Our study should aid managers in avoiding what appears to be a common misconception, i.e., that simply hiring marketing analysts who know how to perform marketing analytics will be sufficient for a firm to benefit from marketing analytics. Instead, we find that TMT involvement and a suitable analytics culture that supports the use of marketing analytics (along with the appropriate IT and data infrastructure) are necessary for the firm to see the benefits of greater deployment.
5.3. Limitations and further research

Although we believe that we have broken new ground with this work, there are clear limitations, several of which provide avenues for future research. First, while our robustness analysis shows that the effects that we report are associated with financial returns, our main measures are attitudinal, not objective. In addition, we do not examine the actual return that a firm could expect from its investments in marketing analytics. Thus, obtaining objective data on the costs and benefits that we measure subjectively in this research would be useful.

Second, our findings are correlational, not causal. For example, we find that a higher level of analytics skills and culture ceteris paribus is associated with the deployment of analytics, which in turn, is associated with higher firm performance. However, we cannot make causal claims regarding these relationships. Future research could be based on longitudinal data for a sample of firms to track changes in the precursors of the deployment of marketing analytics.

### Table 8
The effect of analytics deployment on (objective) firm performance (=DV).

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Model 1: Objective ROA (Time 1) Parameter estimate</th>
<th>Model 1: Objective ROA (Time 1) t-Value</th>
<th>Model 2: Objective ROA (Time 2) Parameter estimate</th>
<th>Model 2: Objective ROA (Time 2) t-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deployment of Analytics</td>
<td>.45**</td>
<td>3.06</td>
<td>.24*</td>
<td>2.08</td>
</tr>
<tr>
<td>Needs &amp; Wants Change</td>
<td>.04</td>
<td>.46</td>
<td>.06</td>
<td>.83</td>
</tr>
<tr>
<td>Competition</td>
<td>.11</td>
<td>1.09</td>
<td>.10</td>
<td>1.26</td>
</tr>
<tr>
<td>Analytics Prevalence</td>
<td>.08</td>
<td>.87</td>
<td>.11</td>
<td>1.43</td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depl × Competition</td>
<td>.12</td>
<td>1.60</td>
<td>.11†</td>
<td>1.79</td>
</tr>
<tr>
<td>Depl × Needs &amp; Wants Change</td>
<td>.13*</td>
<td>2.15</td>
<td>.13**</td>
<td>2.68</td>
</tr>
<tr>
<td>Depl × Prevalence</td>
<td>.03</td>
<td>.46</td>
<td>-.04</td>
<td>-.63</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.00</td>
<td>29.14</td>
<td>4.75</td>
<td>35.58</td>
</tr>
<tr>
<td>R²</td>
<td>32.5%</td>
<td>36.3%</td>
<td>28.9%</td>
<td>32.5%</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>24.7%</td>
<td>28.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-value (7,60)</td>
<td>4.14</td>
<td>4.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-probability</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For ease of interpretation, we mean-centered the focal variables (i.e., deployment of analytics, needs and wants change, competition, and analytics prevalence) before creating the interaction terms (Echambadi & Hess, 2007). **t ≥ 2.576, p < .01; *t ≥ 1.96, p < .05; †t ≥ 1.645, p < .10.

---

![Fig. 4. Structural equation model results using objective ROA (Time 2) as performance measure. Overall, the model fits the data reasonably well; $\chi^2 = 149.744;$ CFI = .932; RMSEA = .089; 90% confidence interval of RMSEA = [.060, .117]. ***$t$ ≥ 3.291, p < .001; **$t$ ≥ 2.576, p < .01; *$t$ ≥ 1.96, p < .05; †$t$ ≥ 1.645, p < .10.](image1)

![Fig. 5. Contemporary vs. carryover effects on firm performance. This linear combination analysis shows that the highest Adj. R² occurs for $\lambda = .4$. This result suggests that the deployment to performance link is strongest with an objective performance variable that gives 40% of the weight ($\lambda = .4$) to contemporary effects on firm performance and 60% to carryover effects.](image2)
### Table 9: Robustness of the deployment to performance link.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter estimates and significance levels (two-sided)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H1: Deployment of analytics is positively correlated with performance measure ($p &lt; .05$)</td>
<td>H2: The interaction between deployment of analytics and competition is significant ($p &lt; .05$)</td>
</tr>
<tr>
<td>OLS regression using objective performance measure (ROA) at Time 1</td>
<td>β Depl. of analytics = .45; $p = .003$</td>
<td>✓</td>
</tr>
<tr>
<td>SEM in which we averaged the responses of multiple respondents of each firm</td>
<td>β Depl. of analytics = .093; $p = .18$</td>
<td>✓</td>
</tr>
<tr>
<td>OLS regression using objective performance measure (ROA) at Time 1 and including quadratic term of deployment of analytics</td>
<td>β Depl. of analytics = .34; $p = .016$</td>
<td>✓</td>
</tr>
<tr>
<td>OLS regression using objective performance measure (ROA) at Time 1 and including control variables</td>
<td>β Depl. of analytics = .42; $p = .008$</td>
<td>✓</td>
</tr>
<tr>
<td>OLS regression using objective performance measure (ROA) at Time 2</td>
<td>β Depl. of analytics = .17; $p = .111$</td>
<td>✓</td>
</tr>
<tr>
<td>Mixture regression model (one-class model)</td>
<td>β Depl. of analytics = .17; $p = .116$</td>
<td>✓</td>
</tr>
<tr>
<td>SEM using single-item sales growth measure from survey instrument</td>
<td>β Depl. of analytics = .171; $p = .016$</td>
<td>✓</td>
</tr>
<tr>
<td>SEM using profit and ROI measures from survey instrument</td>
<td>β Depl. of analytics = .198; $p = .004$</td>
<td>✓</td>
</tr>
</tbody>
</table>
analytics to determine how they affect deployment and how changes in deployment affect firm performance. Such research should be feasible because many firms are still in the early stages of deploying marketing analytics.

Third, our results are based on the overall deployment and impact of marketing analytics. Additional research is needed to understand the performance implications associated with different types of analytics (e.g., embedded automated models vs. interactive decision support), as well as from various aspects of analytics implementation, such as the nature of the decisions/actions supported by analytics (e.g., segmentation, targeting, forecasting, pricing, sales), and the penetration of marketing analytics into non-marketing decisions and actions.

Fourth, our results are based on and limited to very large U.S. firms. Extending this work to other geographies and to the much larger universe of medium-sized and small firms would be useful.

Despite these limitations, we believe that beyond their theoretical interest, our framework and findings should prove useful for managers who are seeking a framework that will aid them in deploying their marketing analytics investments most effectively. Our results also provide a bit of a cautionary tale: Without TMT advocacy and support, the necessary investments in data, analytic skills, and a supportive analytics culture are unlikely to occur. We hope that the modest step we have taken here to address the performance implications of marketing analytics will prove provocative and spawn additional research in this important area.

### Appendix A. Scale Items

<table>
<thead>
<tr>
<th>Measure</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top management team advocacy</td>
<td>1. Our top management has a favorable attitude towards marketing analytics.</td>
</tr>
<tr>
<td>α = .84</td>
<td>2. Our annual reports and other publications highlight our use of analytics as a core competitive advantage.</td>
</tr>
<tr>
<td>Average variance extracted (AVE) = 0.659</td>
<td>3. Our top management expects quantitative analysis to support important marketing decisions.</td>
</tr>
<tr>
<td>Analytics culture</td>
<td>4. If we reduce our marketing analytics activities, our UNIT’s profits will suffer.</td>
</tr>
<tr>
<td>α = .87</td>
<td>5. We are confident that the use of marketing analytics improves our ability to satisfy our customers.</td>
</tr>
<tr>
<td>AVE = 0.692</td>
<td>6. Most people in my unit are skeptical of any kind of analytics-based results (R).</td>
</tr>
<tr>
<td>Marketing analytics skills</td>
<td>7. Our people are very good at identifying and employing the appropriate marketing analysis tool given the problem at hand.</td>
</tr>
<tr>
<td>α = .90</td>
<td>8. Our people master many different quantitative marketing analysis tools and techniques.</td>
</tr>
<tr>
<td>AVE = 0.777</td>
<td>9. Our people can be considered as experts in marketing analytics.</td>
</tr>
<tr>
<td>Data and IT</td>
<td>10. We have a state-of-art IT infrastructure.</td>
</tr>
<tr>
<td>α = .72</td>
<td>11. We use IT to gain a competitive advantage.</td>
</tr>
<tr>
<td>AVE = 0.903</td>
<td>12. In general, we collect more data than our primary competitors.</td>
</tr>
<tr>
<td>Deployment of analytics</td>
<td>13. Virtually everyone in our UNIT uses analytics based insights to support decisions.</td>
</tr>
<tr>
<td>α = .82</td>
<td>14. In our strategy meetings, we back arguments with analytics based facts.</td>
</tr>
<tr>
<td>AVE = 0.657</td>
<td>15. We regularly use analytics to support decisions in the following areas (average score across 12 areas to choose from: pricing, promotion and discount management, sales-force planning, segmentation, targeting, product positioning, developing annual budgets, advertising, marketing mix allocation, new product development, long-term strategic planning, sales forecasting + 2 open ended areas).</td>
</tr>
<tr>
<td>Firm performance</td>
<td>Please circle the number that most accurately describes the performance of your UNIT in the following areas relative to your average competitor (1 = well below our competition; 7 = well above our competition) Please consider the immediate past year in responding to these items.</td>
</tr>
<tr>
<td>α = .81</td>
<td>16. Total Sales Growth.</td>
</tr>
<tr>
<td>AVE = 0.639</td>
<td>17. Profit.</td>
</tr>
<tr>
<td></td>
<td>18. Return on Investment.</td>
</tr>
<tr>
<td>Competition</td>
<td>19. We face intense competition.</td>
</tr>
<tr>
<td>Needs and wants change</td>
<td>20. Our customers are fickle—their needs and wants change frequently.</td>
</tr>
<tr>
<td>Industry prevalence</td>
<td>21. Marketing analytics are used extensively in our industry.</td>
</tr>
</tbody>
</table>

### References


Uncovering audience preferences for concert features from single-ticket sales with a factor-analytic random-coefficients model

Wagner A. Kamakura a,⁎, Carl W. Schimmel b,1

a Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708, United States
b School of Music, Illinois State University, Campus Box 5600, Normal, IL 61790-5600, United States

A R T I C L E   I N F O

Article history:
First received in 17, November 2011 and was under review for 6½ month
Available online 24 October 2012
Area Editor: Harald J. Van Heerde

A B S T R A C T

To better plan their programs, producers of performing arts events require forecasting models that relate ticket sales to the multiple features of a program. The framework we develop, test and implement uncovers audience preferences for the features of an event program from single-ticket sales while accounting for interactions among program features and for preference heterogeneity across markets. We develop a factor-analytic random-coefficients model that overcomes four major methodological challenges. First, the historical data available from each market is limited, preventing the estimation of models at the market level and requiring some form of shrinkage estimator that also takes into account the diversity in preferences across markets as well as the fact that preferences for the many (26 in our application) program features are correlated across markets, requiring the estimation of a large covariance matrix for these preferences across markets. Our proposed factor-analytic regression formulation parsimoniously captures the principal components of the correlated preferences and provides shrinkage estimates at the individual market level. The second challenge we face is the fact that orchestras differ on how they sell season subscriptions, leading to substantial unobserved effects on ticket sales across orchestras; an added benefit of our random-coefficients approach is that it incorporates a random effect that captures any shift in the dependent variable caused by unobservable factors across all events in each individual market, such as the unobservable effect of season subscriptions on single-ticket sales. The third methodological challenge is that program features are likely to interact requiring the estimation of a large set of pair-wise interactions. We solve this problem by mapping the interactions on a reduced space, arriving at a more parsimonious model formulation. The fourth methodological challenge relates to implementation of the model results beyond the relatively small sample of markets for which historical data were available. To overcome this limitation, we demonstrate how our model can be applied to markets not included in our sample, first using only managerial insight regarding the similarity between the focal market and the ones in our sample and by updating this subjective prior as ticket sales data become available.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In 2007, not-for-profit performing arts organizations in the United States garnered revenues of $5.6 billion, of which private and government contributions accounted for approximately two-thirds (NFAH, 2011). Highly dependent upon governmental and corporate support, these organizations must balance adherence to their mission statements with a difficult financial reality. Music ensembles, opera companies, independent theater companies, and dance companies must plan their performance offerings to meet organizational goals, such as providing community education, continuing a classical tradition, and promoting a new repertoire. At the same time, their “product” must satisfy attendees as well as donors. Managers of these organizations must sometimes consider demand when making programming decisions to remain financially viable. Attention to demand is becoming even more important as financial support from government and corporations become scarcer.

Symphony orchestras offer a typical example. As in other not-for-profit arts organizations, ticket revenue accounts for approximately one-third of an average symphony orchestra’s gross receipts (LOAO, 2001-2006). Solid ticket sales are therefore a crucial component of any orchestra’s financial health. However, assembling a concert season only from popular standard repertoire and film score excerpts would be contrary to the mission statements of most orchestras. The use of statistical methods to more accurately determine the financial ramifications of programming decisions would provide a way of making wiser decisions regarding the concert revenue portion of the budget, enabling music directors to make more adventurous programming choices. Several studies have confirmed the effectiveness of quantitative techniques when applied...
to forecasting for performing arts events (e.g., Bennett, 2002; Weinberg and Schachmut, 1978). Yet, American orchestras are far behind the for-profit sector in adopting analytical techniques for optimizing ticket sales. One reason for this fact may be that orchestra marketing directors are sometimes lacking in the sufficient musical background to plan a financially viable season, while music directors’ programming decisions are often informed more by a sense of musical heritage than by any true awareness of the subtleties of audience preferences. Orchestra boards are understandably reluctant to treat “high art” like a consumer product because a slavish attention to market factors would be antithetical to the mission statements of these organizations. However, when implemented in full consideration of these artistic aims, certain methods of econometric modeling could enhance the financial stability of these organizations. As Stephen Belth, Executive Director of the Long Island Philharmonic explains, “If one accepts that concerts should be well attended, and in light of current audience patterns, orchestra organizations must explore every means available to improve long-term attendance of the concert-going public” (Belth, 1999).

Although the task of an orchestra’s marketing director is to promote concert events as they are planned by the music director, in actuality, programming decisions are not made by the music director alone. If forecasting tools were made available to a marketing director, perhaps incorporating the orchestra’s own historical data, he or she could play a more relevant role in the programming process, helping to fine-tune the season’s offerings to achieve a balance between the orchestra’s educational and cultural obligations to its community and its own financial wellbeing. One of the roles of the executive director should be to facilitate the inclusion of the marketing director’s contributions in the music director’s final programming decisions to enhance program appeal to the target audience.

The performing arts have received limited attention in the forecasting literature. Past efforts to build ticket sales forecasting models for performing arts organizations are relatively few in number, and no sources were found in the literature that attempted to model ticket sales for orchestras in particular. Weinberg and Schachmut (1978) developed a practical planning model called “Arts Plan” that used dummy variable regression analysis to predict attendance at a series of music and dance events by taking into account seasonality, product variation, day of performance, number of performances, popularity of offering, and year. In a subsequent article, Weinberg (1986) published a revision of the Arts Plan model which addressed how accuracy might be improved by a manager’s knowledge-based modification to a forecast that was based solely on historical data.

In their study modeling Broadway show attendance, Reddy, Swaminathan, and Motley (1998) reported the significance of newspaper reviews, previews, advertising, high-profile actors, and longevity. They found, however, that attendance was relatively price inelastic, echoing similar findings by previous studies of orchestra attendance (e.g., Currim, Weinberg, & Wittink, 1981). Putler and Lele’s model (2003) considered market size and venue capacity, and included subjective variables for the number of potential customers and promotional effort, as well as dummy variables for features such as day of week, season, the existence of a competing event, and the “difficulty” of the presentation. In the application of their model to a series of theater productions, the marketing director used model estimates to adjust her own expectations of how each production should be marketed and scheduled. Their model takes into account the number of events in a “run” as well as the possibility of a “sellout,” a factor that has not been taken into account in other forecasting models for the performing arts.

Our study differs from previous forecasting models for the performing arts in two important ways. First, our main focus is on understanding how audiences from different orchestral performances respond to the many features of the program for each of multiple concerts, along with their interactions. While previous forecasting models for the performing arts have considered some program features such as show type, lead performer (Reddy et al., 1998), play type, plot complexity, and social issues addressed (Putler & Lele, 2003), these studies focused on the main effects of a limited number of program features on a single audience or market. In contrast, our proposed model considers potential synergies or negative interactions among the multiple features of a concert program, while also accounting for the fact that different orchestras cater to different audiences, possibly with distinct preferences for certain features of a concert program. Second, because our main interest is in understanding how audiences in different markets react to multiple features of concert programs, we will investigate audience preferences as indicated by single-ticket sales.

Our task presents four major methodological challenges. First, most symphony orchestras in the US have limited historical concert data, which prevents the reliable construction of statistical models at the orchestra level. Therefore, one must leverage available data from all orchestras in the US market. However, the audiences and their preferences for different orchestras are quite distinct, thus requiring a model that accounts for unobserved heterogeneity across audiences using a random-coefficients formulation (Swamy, 1970), a compromise between aggregate and orchestra-level estimation. Unfortunately, preferences for the many (26 in our application) program features are most likely correlated across audiences, requiring the estimation of a large covariance matrix, which would render the typical econometric model over-parameterized for the limited data available. We tackle this particular challenge by capturing the principal components of this large covariance matrix with a formulation that is parsimonious, and therefore more robust for predictions.

Another methodological challenge in uncovering programming preferences for different audiences from ticket sales is the fact that concerts are attended by both single ticket buyers and season subscribers. Furthermore, for many of the orchestras, subscriptions are either unavailable or cannot be directly apportioned to individual events because season subscribers either buy pre-set bundles of events or are granted access to a certain number of events. The fact that orchestras differ in their reliance on subscription sales is likely to create a bias in the estimates of audience preference for program features, unless these (and other) systematic differences across orchestras are taken into account. Our model framework accounts for differences across orchestras through a random-effects formulation which captures the impact of unobservable factors on each market’s sales response, such as the fact that these markets differ on the capacity of their venues, and on the percentage of total venue occupancy represented by single-ticket sales.

The third methodological challenge we overcome with our proposed framework is that program features are likely to interact, so that pairing certain program features leads to a concert that is more (or less) attractive than the mere sum of its parts. For example, a guest soloist who is known for his interpretations of the Rachmaninoff Piano Concertos would attract more spectators when performing Rachmaninoff rather than Mozart. The bundling of program features such as soloists, repertoire, and conductor is of great concern to music and marketing directors. Interaction effects have largely been ignored in previous performing arts forecasting models. These feature interactions pose a similar methodological challenge as discussed above because they would require the estimation of a large set (18 × 18/2 = 171 in our application) of pair-wise interactions, rendering the model over-parameterized and thereby less stable for forecasting purposes. We solve this problem by mapping the interactions on a reduced space, arriving at a more parsimonious model formulation.

The final methodological challenge relates to implementation of the model results. Our sample covers only 47 venues for symphony orchestras. While this sample is representative of symphony orchestras in the US, it includes less than 15% of all professional orchestras in the US (LOAO, 2010), which would limit the direct application of our empirical results to only these 47 venues. In other words, while
our results are individually valid for each of the sampled orchestras (as we demonstrate in our predictive validity tests) and generalizable for the population of orchestras, a program director for an orchestra not included in our sample would also want to benefit from insights regarding its own audience, when planning a concert program for his/her orchestra, which requires extrapolation of these results beyond the sample used in our study. To overcome this limitation, we demonstrate how our model can be applied by the music and/or marketing director of an orchestra not included in our sample, first by using managerial insight regarding the similarity between his/her orchestra and the ones included in our sample and later, by updating this subjective estimation as ticket sales data become available for new concerts performed by his/her particular orchestra. With this valuable feature, our model can be used to produce customized insights for other orchestras beyond those we analyzed, as long as its managers can see some similarity between their market and the 47 markets we studied, and/or have sales and program data to tailor our model to their particular market.

In the next section we develop and describe a model that attempts to overcome these four main methodological challenges in uncovering audience preferences revealed through ticket sales. This is followed by a description of our data and the application of the proposed model to them, along with a discussion of the results, predictive validity tests, and implementation issues.

1.1. Modeling audience preferences for program features revealed through concert attendance

Our primary goal in developing our model is to use ticket sales to uncover audience preferences for program features so the results can aid orchestra music and marketing directors in the development of concert programs that achieve their cultural and marketing goals. We fit the model using single-ticket sales as an indicator of audience preferences and test its predictive validity on hold-out forecasts.

As discussed previously, the sample data available from many of the orchestras are limited relative to the large number of features (and pairwise interactions) that must be considered to render the model managerially useful. Moreover, there are several unobservable factors that are idiosyncratic to specific audiences/markets (e.g., season subscriptions, venue size and market potential) that shift observed sales and could lead to biased aggregate estimates of audience response to concert features. For this reason, we must reach a compromise between aggregate (across orchestras) and individual (specific to each orchestra) estimation, which can be achieved with a random-coefficients regression (Swamy, 1970). However, a simple random-coefficients regression that assumes independent random coefficients would be overly simplistic because preferences for program features are most likely correlated across markets, and might also be correlated with unobservable orchestra effects that are captured by the random intercept. On the other hand, as discussed earlier, a full-covariance model would require the estimation of a large number of parameters for the covariance of the random coefficients alone. Again, we reach another compromise between model flexibility and parsimony by capturing the first P principal components (determined empirically) of the covariance among the random regression coefficients (including the intercept), which leads to a more parsimonious and robust formulation. With these considerations in mind, we define our regression model as

\[ y_{it} = \sum_k \left( \beta_k + \sum_{p=1}^{P} \lambda_{kp} x_{ip} \right) x_{ikt} + \sum_k \sum_{k' \neq k} \theta_{kk'} x_{ikt} x_{ik't} + e_{it} \]  

where:

- \( y_{it} \) = ticket sales on concert t by orchestra i
- \( x_{ikt} \) = feature k of the program for concert t by orchestra i, which includes a unit vector to capture orchestra-level intercepts
- \( \beta_k \) = average (across orchestras) regression coefficient for feature k (including intercepts, estimated as random effects)
- \( x_{ip} \) = score of orchestra i on latent dimension p, independent distributed \( N(0,1) \)
- \( \lambda_{kp} \) = weight for feature k on latent dimension p
- \( \theta_{kk'} \) = coefficient capturing the pair-wise interaction between program features k and k'
- \( e_{it} \) = independent distributed random error \( \sim N(0,\sigma) \)

The first term in the right-hand side of Eq. (1) shows the principal-components decomposition of the random regression coefficients, where the response coefficient for a specific orchestra i on a program feature k is given by \( \beta_k + \sum_{p=1}^{P} \lambda_{kp} x_{ip} \). Note that these random coefficients include a random-effect (intercept) coefficient for each orchestra, which captures the effects of all factors affecting sales that are unobservable at the audience/market level, such as the fact that orchestras differ on the unobservable way they handle season subscriptions. Because the latent scores z are independent standardized normals, the covariance between the random regression coefficients for features k and k' is given by \( \sum_{p=1}^{P} \lambda_{kp} \lambda_{kp'} \), thereby allowing us to capture K(1+1)/2 covariance terms with a set of K*P parameters and maintain parsimony as long as the number of dimensions P is smaller than the number of program features K. Notice that we allow for heterogeneity across markets not only in their response to the program features but also in the intercept. This heterogeneity in the intercept is critical to capturing any unobserved effects that are unique to each market. For example, each of the symphony orchestras in our sample (with the exception of four) performs in a single venue, and these venues vary in size across markets. This variation in venue size, along with differences in potential audiences across markets can inject bias into the estimated response coefficients unless they are taken into account, which is accomplished in our model via the random effects.

The second term in the right-hand side of Eq. (1) captures the possible pair-wise interactions among the K program features. As they are specified in Eq. (1) these interactions require the estimation of K(K−1)/2 pair-wise \( \theta_{kk'} \), rendering the model over-parameterized for the available data. For this reason, we use another form of space reduction to map the K(K−1)/2 interactions into a smaller set of K*Q coordinates on Q dimensions (to be determined empirically).

\[ \theta_{kk'} = \sum_{q=1}^{Q} \delta_{kq} \delta_{k'q}, \forall k \neq k', \]  

with \( \lambda_{pq} = 0 \) if \( q > k \) for identification purposes.

Notice that we assume homogeneity in the effect of feature interaction on preferences. We do this for two main reasons. First, we assume that interactions are intrinsic to the concert features, and therefore less affected by preferences. In other words, featuring Yo-Yo Ma playing the piano, rather than the cello would seem odd, therefore less affected by preferences. In other words, featuring Yo-Yo Ma playing the piano, rather than the cello would seem odd, therefore less affected by preferences. In other words, featuring Yo-Yo Ma playing the piano, rather than the cello would seem odd, therefore less affected by preferences. In other words, featuring Yo-Yo Ma playing the piano, rather than the cello would seem odd, therefore less affected by preferences.

1.1.1. Model estimation

Estimation of the proposed model is relatively simple, requiring the combination of the well-known Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) with non-linear least squares estimation in the M-step (for the estimation of the interactions in Eq. (2)). In the M-step, the latent scores z are treated as missing values and replaced by their expectation from the E-step (described next). The model parameters \( \Theta = [\beta, \lambda, \theta, \sigma^2] \) can then
be estimated by minimizing the sum of squared errors across orchestras and concerts

\[\text{SSE} = \sum \sum [y_{it} - \sum \{\beta_k + \sum \{\sum q \delta_{i k q} z_{i k q}\} x_{i d a}\}]^2.\]

which we obtain via standard non-linear least-squares estimation.

In the E-step, the p-dimensional latent scores \(z_i\) for each orchestra \(i\) are obtained, based on the current parameter estimates from the M-step (Eq. (3)), from the following steps:

1. Obtain \(M\) draws of \(p\)-dimensional vectors of independent and identically distributed (i.i.d.) standard normals \(v_{m, i} = 1, 2, \ldots, M\) using a randomized Halton sequence to best approximate the i.i.d. normal distribution (Train, 2003).

2. For each of the \(M\) draws \(v_{m, i}\), compute the conditional likelihood for each observation \((y_{it}, x_{i})\). Given the current parameter estimates \(\Theta = [\beta, \lambda, \delta, \sigma^2]\), by replacing the unknown heterogeneity factor scores \(z_i\) by \(v_{m, i}\),

\[L(Y_{it}, X_{it} | \Theta, v_m) = \prod \prod \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2\sigma^2} v_{im}^2\right]\]

where \(v_{im} = y_{it} - \sum (\beta_k + \sum \{\sum q \delta_{i k q} z_{i k q}\} X_{i d a}) - \sum \sum \{\sum q \delta_{i k q} z_{i k q}\} X_{i d a} X_{i d a}\)

and \(v_{mp}\) is one of the i.i.d. random draws on dimension \(p\).

3. Compute new estimates of the heterogeneity and non-stationary factor scores

\[z_{i} = \frac{\sum_{m=1}^{M} [v_{im} \cdot L(Y_{it}, X_{it} | \Theta, v_m)]}{\sum_{m=1}^{M} L(Y_{it}, X_{it} | \Theta, v_m)}\]

The expectation and maximization steps described above are repeated until the parameters from the current M-step are reasonably close to those obtained in the previous cycle. The dimensions of the two reduced spaces (\(P\), capturing heterogeneity in preferences across markets, and \(Q\), capturing interactions among program features) are determined empirically by fitting the model with different values of \(P\) and \(Q\) and comparing their expected predictive fits via the Bayesian Information Criterion (Schwarz, 1978) or Akaike’s Consistent Information Criterion.

1.1.2. Implications of the model

Because the proposed model is basically a random-coefficients regression, albeit with a factor structure in the random coefficients and with homogeneous feature interactions, it produces estimates both at the aggregate and orchestra levels. The parameter \(\beta_k\) captures the average preference across all markets for the program feature \(k\), including the intercept. However, the model also accounts for the fact that program features interact, allowing for the possibility that combinations of program features might produce results that are greater than the individual contributions of each feature, while other combinations might interact negatively. The total aggregate appeal of a particular program, as defined by the features \(X = \{x_k, k = 1, 2, \ldots, K\}\), can be computed from the model estimates \((B, \Delta)\) as

\[V(X, B, \Delta) = \sum \{\beta_k x_k + \sum \{\sum q \delta_{i k q} z_{i k q}\} x_{i d a}\}.\]

Most importantly, if the music or marketing director for a specific orchestra \(i\) with factor scores \(z_i\) (obtained during model calibration) wants to know how the same program might work on her particular audience, the audience-specific appeal can be computed as

\[V(X, B, \Delta, A, Z_i) = V(X, B, \Delta) + \sum \{\sum q \delta_{i k q} z_{i k q}\} x_{i d a}.\]

Note that the results in Eq. (5) pertain to a specific audience for orchestra \(i\), thereby taking into account the idiosyncratic preferences in that particular market, but utilizing information from all other markets through shrinkage estimation, which is accomplished by the factor loadings \(\lambda\) estimated across all markets and the factor scores \(z_i\) obtained for that specific market.

It is important to note that the audience/market-level effects in Eq. (5) include an intercept for market/audience \(i\), which captures the effects of unobservable factors affecting ticket sales for all events for market/audience \(i\). Such factors include the fact that orchestras differ in the (unobservable) way they handle season subscriptions, their venue capacity, and market potential. These unobservable differences across orchestras are captured in the random intercept which, in our factor-analytic formulation, is also allowed to correlate with the observable features of the concerts offered by the orchestra.

We believe this feature of the proposed model is particularly important because problems in measuring the dependent variable lead to unobservable differences across orchestras, which must be taken into account for a better assessment of audience preferences. Our proposed formulation accounts for these unobservable differences through random intercepts across orchestras.

2. Empirical illustration

We illustrate and test our proposed model using data collected for 47 symphony orchestra markets in the United States. These orchestras were among those included in the American Symphony Orchestra League’s 2004–2005 Orchestral Repertoire Report. At that time, only orchestras with budgets of more than $1.7 million were included. No subscription ticket data were included, nor were data for complimentary or reduced-rate tickets. Because subscription ticket sales are handled in different ways by each orchestra and are not easily assigned to individual events (and consequently to programming decisions for each concert), our model focuses solely on single-ticket sales. Although orchestras are traditionally dependent on a strong subscriber base, this focus on single-ticket sales also responds to a recent trend toward higher single-ticket sales and lower season subscriptions. Furthermore, some scholars (Johnson & Garbarino, 2001) have argued that the costs of retaining and renewing subscriptions outpace the transaction costs of single tickets, and that successful organizations with too many subscribers may lose revenue on seats that are not sold for the optimal price.

Obviously, because orchestras differ on their reliance on season subscriptions, omission of this and other orchestra-level factors will shift the dependent variable for each orchestra in an unobservable fashion, leading to biased estimates of the sales–response model unless these potential biases are accounted for. However, as described previously (Eq. (5)), our modeling framework directly incorporates adjustments for unobservable differences across orchestras via a random-effects formulation, thereby accounting for these unobserved disparities between orchestras. One especially important and useful feature, given the substantial unobserved differences across markets, is the fact that the proposed modeling framework accounts for these unobservable
effects through random-effects across orchestras (which are also allowed to correlate with responses to observable concert features).

Participating orchestras provided all available single-ticket data, usually several seasons’ worth, ranging from 23 concerts on one extreme to 619 concerts at the other, with an average of 110 concerts per orchestra. We also collected data from one orchestra that provided only 15 events, which we used for an out-of-sample illustration, further demonstrating how the results from our sample of 47 orchestras can be generalized for an orchestra not included in our sample.

The dependent variable for our model is log(occupancy), where occupancy is defined as the number of single tickets divided by the venue capacity. We chose to define our dependent variable as venue-occupancy attributed to single tickets for two main reasons. First, almost all the orchestras in our sample performed in a single venue. Four exceptions performed in two venues, in which case the secondary venue was reserved for distinct programs, such as chamber music. Each of these four exceptions was split into two unique markets. Because each orchestra is uniquely associated with a single venue, differences in venue size are directly captured by the random effect, as discussed earlier, thereby justifying the use of occupancy as the dependent variable. Second, we did not have access to the number or proportion of seats taken by season subscribers in each concert, which we believe varies across orchestras. This is also captured by the random intercepts because we define our dependent variable as the proportion of seats occupied by single-ticket buyers. Although it would have been desirable to include the number of season tickets assigned to each concert in our dependent variable, this information, along with complimentary or reduced-price tickets, was not available to us. We see this lack not only as a data limitation but also as a practical problem that makes our random-effects formulation particularly valuable in this illustration because it accounts for unobservable differences across orchestras.

We constructed the following dummy variables for use as predictors in the model:

<table>
<thead>
<tr>
<th>Event Feature</th>
<th>Dummy Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturday evening</td>
<td>Choral (other than popular)</td>
</tr>
<tr>
<td>Sunday afternoon</td>
<td>Mozart (other than popular)</td>
</tr>
<tr>
<td>Sunday–Thursday evening</td>
<td>Beethoven (other than popular)</td>
</tr>
<tr>
<td>January, February, March</td>
<td>Tchaikovsky (other than popular)</td>
</tr>
<tr>
<td>April, May, June</td>
<td>Brahms (other than popular)</td>
</tr>
<tr>
<td>September, October</td>
<td>R. Strauss</td>
</tr>
<tr>
<td>Star Soloan</td>
<td>Prokofiev</td>
</tr>
<tr>
<td>Soloist with 20 or more appearances in sampled</td>
<td>Other Works (pre-Romantic)</td>
</tr>
<tr>
<td>concerts</td>
<td>Other Works (Romantic)</td>
</tr>
<tr>
<td>Guest conductor</td>
<td>Other Works (late Romantic)</td>
</tr>
<tr>
<td>Includes at least one “most popular” piece</td>
<td>Other Works (turn of the century)</td>
</tr>
<tr>
<td></td>
<td>Other Works (twentieth century)</td>
</tr>
<tr>
<td></td>
<td>Contemporary (alive)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on discussions with orchestra marketing directors, we ascertained that seasonality was a relevant factor in ticket sales. To preserve parsimony, we categorized each concert into one of four “seasons” rather than into months with November–December used as the basis. To compile a list of star soloists, we aggregated the Billboard data from 1977 to 2005 (the top fifteen albums of each month’s report). Soloists who appeared more than 6 times in the Top Classical Albums (Billboard, 2005a) charts in the twenty years prior to a given concert were considered “star soloists,” as were soloists who appeared more than 6 times in the Top Crossover Classical Albums (Billboard, 2005b) charts in the ten years prior to a given concert. Because some famous soloists focus more on live performances than on recording, we also included a variable to indicate if a concert program featured a soloist who made 20 or more “appearances” in the sampled data. Some conductors are especially popular and draw large audiences when they are in town (e.g., Michael Tilson Thomas in San Francisco); conversely, some conductors are famous enough that they will attract a large audience when they guest conduct (e.g., most recently Gustavo Dudamel). We attempted to capture this effect with a “guest conductor” variable.

The works being performed play a large role in most concertgoers’ decisions to attend (John & Knight Foundation, 2003, Table B1). Because only 6% of potential classical consumers consider themselves “very knowledgeable” about classical music (50% call themselves “not very knowledgeable”), and because concerts with familiar music are preferred, it follows that there is a small collection of musical works that would significantly increase revenue and attendance if included on a program. We expected that concert programs which include better-known works would have a strong tendency to attract bigger audiences and that the “popularity” of a given work would have a stronger positive correlation with ticket revenue than any other repertoire-related criterion (e.g., nationality, composer, genre). For this reason we established the categories “most popular” and “other popular.” Because six in ten orchestra ticket buyers listen to classical music on the radio daily or several times a week, we used a review of eight “classical countdown” lists assembled by various radio stations nationwide (as of April 2006) to determine the prevailing tastes of concertgoers. Works that were in the top 28 of all eight lists were deemed “most popular,” and works which were in the top 28 of at least two (but no more than seven) of the lists were assigned to the “other popular” category. Choral works are known to draw bigger audiences, even if they were not “popular,” so we established a separate choral work category. Works in this category did not include “popular” works, which were already slotted to the “most popular” and “other popular” categories.

In a recent study, of the 43% of potential classical consumers who indicated that they had a favorite composer, 23% named Beethoven, 23% Mozart, 15% Bach, 9% Tchaikovsky, and 4% Chopin (John & Knight Foundation, 2003). While the tastes of actual concertgoers are almost certainly more diverse, these data suggest that the inclusion of music by certain renowned composers on a particular program would increase both attendance and revenue. This is also substantiated by the fact that people tend to prefer music with which they are somewhat familiar (John & Knight Foundation, 2003). Therefore, because the works of Mozart, Beethoven, Tchaikovsky, Brahms, Richard Strauss, and Prokofiev were sufficiently represented in the data, their works were assigned their own dummy variables. Popular works, however, were excluded from these categories (for example, for the purposes of our study, Beethoven’s Ninth Symphony was considered a “Most Popular” work rather than a work by Beethoven). Works by other composers were categorized by era. Pre-Romantic composers include those born before 1800, Romantic composers include those born between 1800 and 1849, Late Romantic composers include those born between 1850 and 1873, Turn-of-the-Century composers include those born between 1873 and 1899, Twentieth Century composers include composers who were born in the Twentieth Century who are no longer alive, and Contemporary (“Alive”) composers include all living composers. Several less commonly found subcategories (music of Igor Stravinsky and Maurice Ravel, and music of Twentieth Century American composers) are omitted from the categorization and comprise the default value. Table 1 presents summary statistics for all the variables used in our study. We constructed a “Quality-Adjusted Price” measure from the inflation-adjusted average ticket price for each concert. This calculation was performed to account for the fact that prices might be endogenously determined based on program costs. For example, a program featuring a star soloist such as Yo-Yo Ma or a commissioned piece by a contemporary composer will have higher costs, and might also be more appealing to the audience. Therefore, the marketing directors may wish to recover the higher costs with higher than average ticket prices for the event.

To the best of our knowledge, this potential endogeneity has been ignored in the past marketing literature in the performing arts, which may explain the low price elasticities reported in those studies. To
adjust prices for programming quality, we ran the regression reported in Table 2 across all events and for all orchestras, and utilized the regression residuals as the quality-adjusted prices, assuming that the cost structure for these program features is common across orchestras. Notice that we only adjusted prices for features that are directly related to program content, while the sales model (to be discussed later) was estimated on all event features, which ensures the identification of both models. As one would expect, star soloist and soloist with more than 20 concerts show a positive and statistically significant coefficient indicating that these features tend to be offered at higher prices across orchestras. Featuring a guest conductor was not found to positively affect ticket prices, while the inclusion of popular pieces tended to have a negative effect on ticket prices, most likely because these pieces are associated with regular events sold at normal (lower) average ticket prices. Most importantly, the residuals from this hedonic price regression provide a more valid measure of ticket prices that is adjusted for the extra cost associated with special features.

Another potential form of endogeneity or source of supply-side effects is the possibility that program/music directors might build and schedule programs to best match the expected audiences to optimally occupy their venue. Although this calculation is possible in theory, requiring a simultaneous supply/demand model across all orchestras and concerts with the proper instruments to disentangle the two sides, in practice, program/music directors operate under numerous unobservable (to the researcher) constraints. For example, some orchestras do not own their venues, in which case venue availability is of primary importance. The availability of an orchestra’s music director is also paramount. In some cases, orchestras need to alternate their regular season Classical subscription series with a Pops series, which means that the availability of the Pops conductor and the Pops performers must be taken into account as well. Smaller and mid-size orchestras are often required to contract their performers on a weekly or per-service basis, further limiting scheduling flexibility. In many smaller and mid-size markets, orchestral players perform and rehearse with other orchestras in neighboring cities (e.g., Portland, ME and Providence, RI). Particular star soloists who are in demand may or may not be available on a particular date, and their inclusion in the calendar will also affect the other components of the program. Given these and other complexities, as well as the practical difficulties in finding proper and relevant instruments, we leave a more comprehensive demand/supply equilibrium model for future research. It should be noted, however, that we estimate individual orchestra effects via the random-coefficients formulation, which captures any endogeneity bias at the orchestra level.3

2.0.1. Empirical results

To test for the predictive validity of our proposed model, we held out data from the last five concerts from each of the 47 orchestras in our calibration sample so we could compare the forecasts with actual ticket sales for these five future concerts in true step-ahead forecasts. To demonstrate how the results can be generalized beyond our sample of orchestras, we held out the data from one of the smaller orchestras, so that the results obtained from the other 47 orchestras could be used to produce out-of-sample predictions for the orchestra that was excluded from the calibration sample.

As we discussed previously, the proposed model allows for a flexible accounting of heterogeneity in preferences for program features across audiences, and also of the interactions among program features, while maintaining parsimony by capturing the principal components of the covariance matrix of random preferences and by space reduction of the pair-wise interactions. However, this parsimony is determined by the choice of dimensions to account for heterogeneity and interactions. To guide our decisions for each dimension, we fit the proposed model under various configurations. Several fit statistics are reported in Table 3.

Table 3 compares the fit for the aggregate model ignoring differences across audiences without interactions (first column) and with interactions mapped onto one or two dimensions (next two columns). As one would expect, there is a small improvement in fit by

3 We also ran a fixed-effects model as an alternative approach to handle endogeneity. We obtained fixed effects that are statistically indistinguishable from the individual-level estimates obtained with our formulation, which supports our model specification.
allowing for feature interactions, indicated by a shift in $R^2$ from 21.2% to 23.9% and 24.6%. This is confirmed by Akaike's Information Criterion (AIC), but not by the Consistent AIC (CAIC), which imposes a more stringent penalty for extra parameters, proportional to the logarithm of the sample size. The improvement in expected predictive performance by allowing for heterogeneity in preferences across audiences is quite clear, particularly for the models with one and two latent factors, in which all indicators favor the more complex formulation. However, the CAIC suggests that the 3-factor solution is over-parameterized, relative to the two-dimensional factor model. In other words, the CAIC criterion indicates that adding more parameters to capture unobserved heterogeneity (i.e., adding more complexity to the covariance matrix of the random coefficients) would cost more, in terms of complexity, than it is worth in terms of predictive fit. In the extreme case, a full specification of the covariance matrix would lead to an even more over-parameterized model (our attempts to estimate this fully saturated model failed to converge).

The AIC criterion suggests that the modeling of feature interactions would improve predictive performance, while the more stringent CAIC does not support the inclusion of these interactions. Because the interactions between program features provide potentially valuable insights to music and marketing directors, we will report the detailed results for the complete model with two latent-factors for audience heterogeneity and two dimensions capturing feature interactions. Later, we will compare the forecasting performance for all versions of the model listed in Table 3.

One potentially important consideration ignored by our model and implemented here, is that if the same type of concert is played over and over again at the same venue, there is likely to be a saturation effect in the sense that attendance will go down. To test the demand-side implication of this phenomenon, we expanded the model with variables representing saturation effects (Van Oest, van Heerde, & Dekimpe, 2010). We did this by adding a lagged binary variable indicating whether the same feature was offered in either of the past two previous concerts in the same venue, creating 18 additional binary predictors for each feature of the concert program (other than scheduling and price). However, this inclusion represented a heavy burden in terms of additional parameters to be estimated (54 additional parameters for a 2-dimensional heterogeneity map, assuming that the saturation effects do not interact). A model comparison in terms of expected predictive validity (measured by CAIC) clearly showed that incorporating the 18 additional saturation predictors led to an over-parameterized model. In other words, the additional predictors did improve fit (as one should expect), but the loss of degrees of freedom was not justified by the improvement in fit, therefore jeopardizing the (predictive) validity of the extended model. We took these results as indication that preference saturation is not a major limitation to attendance in our particular application.

Table 4 shows the parameter estimates for the complete model with two latent factors accounting for heterogeneity in preferences across audiences and two dimensions capturing the interactions among program features, where parameters that are statistically significant at $p<0.01$ are marked in bold characters. For a more intuitive interpretation of the typical effect of program features on single-ticket occupancy, we report in Table 5 the expected average percentage changes in single-ticket occupancy in response to a 10% reduction in quality-adjusted price and with the inclusion of each of the other program features across all orchestras and concerts.

As one would expect, on average (shown in the first column) all markets respond positively to Star Soloist, Most Pop Piece, Other Pop Piece, and Choral, while responding negatively to prices (after correcting for program content). On average, audiences also respond positively to Mozart pieces that are not among the most popular, but respond negatively to less popular works – and works by lesser-known or less popular composers – from the Pre–Romantic, Romantic, Late Romantic and 20th Century periods. For example, the inclusion of masterworks such as Dvořák’s Eighth Symphony and Sibelius’ Violin Concerto, both of which were among the most frequently performed works in the 2005–2006 season, on a program apparently does not translate into ticket sales, perhaps because they are not as well known to the general public.

However, the model suggests that audiences do not respond more negatively to the inclusion of a Contemporary work on the program, than to the inclusion of less-known pieces from the Pre–Romantic or Romantic periods. The reason for this is unclear, but there are several possibilities. Contemporary works can allow for additional marketing possibilities (e.g., the composer is local, the work is about a popular contemporary figure, etc.). Furthermore, recent trends in orchestral composition favor the incorporation of elements from rock and popular music, making contemporary music increasingly accessible to a broader public. And occasionally, film composers – the most famous of contemporary composers – are included on concert programs. Howard Shore’s Lord of the Rings Symphony, for example, was a great financial success in Des Moines and Spokane. On average, audiences also seem to prefer the resident conductor over a Guest Conductor. The number of “star” conductors seems to be low when compared to the number of star soloists; guest conductors are perhaps less well-known to audiences and may be harder to market than the local personality.

The latent factor coefficients (second and third columns of estimates in Table 4), when combined with the factor scores for each orchestra, determine how the preferences of the audience of each orchestra depart from the averages discussed above, while also accounting for systematic differences in single-ticket occupancy across audiences/markets via the intercept. Even though the factor coefficient and respective factor scores were obtained in a very different way (they were obtained as the principal components of the unobservable multivariate distribution of regression coefficients) than standard factor analysis (which is performed directly on observed variables), these estimates are interpreted in a similar way. These factor coefficients are more meaningful when displayed in a vector map, which is shown in Fig. 1, and can be interpreted as factor loadings for the respective regression coefficients. In this map, program features are represented as vector termini pointing to the orchestra positions (defined by the orchestra factor scores) that respond better than average to these program features. For example, the darker vector pointing to intercept indicates that orchestras positioned

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fit statistics for different formulations of the model.</strong></td>
</tr>
<tr>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>Aggregate</td>
</tr>
<tr>
<td>Aggregate + 1D Interactions</td>
</tr>
<tr>
<td>Aggregate + 2D Interactions</td>
</tr>
<tr>
<td>1D Random Coefficients</td>
</tr>
<tr>
<td>2D Random Coefficients</td>
</tr>
<tr>
<td>1D Random Coefficients + 1D Interactions</td>
</tr>
<tr>
<td>2D Random Coefficients + 1D Interactions</td>
</tr>
<tr>
<td>2D Random Coefficients + 2D Interactions</td>
</tr>
</tbody>
</table>

Note: AIC = Akaike Information Criterion.
CAIC = Akaike Consistent Information Criterion.
R2 = % of variance explained in the dependent variable by the model.

---

4 We also tested a simpler and more parsimonious model with a common saturation effect for all program features, but obtained a saturation effect that was not statistically significant at the 0.05 level.
competition from other forms of entertainment in the same market. We see this ability to incorporate unobservable differences across orchestras/markets as a valuable feature of our proposed model for capturing audience response for each orchestra. Most of the program features point in opposite directions along the same line, as indicated by the light-color vector. The program features that seem to distinguish the audiences the most are Prokofiev, Most Pop Piece, Star Soloist, Beethoven and Other Pop Piece in one direction and Brahms, Sun-Thru Eve, Turn of Century and Mozart in the other direction.

Considering that the preference heterogeneity map in Fig. 1 shows the locations where audiences are more/less responsive than average to program features, it would be useful to know where each of the 47 audiences is located in this same map. The locations of the 47 audiences in the same preference heterogeneity map are shown in Fig. 2. Combining Figs. 1 and 2 we can now see that orchestras positioned in the far right tend to occupy their venues with single-ticket sales at a higher rate than other orchestras, while those on the far left have lower than average occupancy with single-ticket sales, again due to many unobservable factors (the remaining seats may be either empty or occupied by season subscribers; we do not know to identify them). Nearly all of the orchestras from large metropolitan areas are located in the right half of Fig. 2; it is reasonable to assume that orchestras in large markets might be better able to fill their venues with single tickets as opposed to season tickets or unsold seats. Furthermore, several of the orchestras shown in the left half of the graph are based in smaller cities but have very large venues and are not likely to fill closer to capacity with single-ticket sales only.

Figs. 1 and 2 suggest that audiences in the top right quadrant of Fig. 2 respond more favorably to Prokofiev, Star Soloist, Most Pop Piece, Other Pop Piece, Beethoven (other than popular), and also buy more single tickets relative to the total capacity of the venue because they are located farther away from the origin along the direction indicated by the intercept vector in Fig. 1. With the possible exception of Prokofiev, these variables represent program features that are more marketable (e.g., “non-popular” Beethoven, which actually has a negative intercept). Orchestras in the top right quadrant may have a more effective non-subscriber marketing strategy or an advantage due to market size. However, several larger orchestras with high single-ticket ratios apparently respond comparatively less favorably to popular program features. These orchestras seen in the lower right quadrant of Fig. 2. One potential explanation for this might be that star soloists and popular classical music offerings are less unique for large markets. The true reason for these correlations with the random intercepts is unknown because these intercepts capture all unobservable factors affecting the general level of single-ticket sales (across events) for each orchestra.

The audience factor scores shown in Fig. 2, combined with the coefficients shown in Fig. 1, provide some insights into the program features that are more likely to produce higher single-ticket sales for each particular audience. For example, CHT and NYP have the largest scores on Factor 1 and therefore are further along in the intercept direction (see Fig. 1), which means that these markets tend to occupy their venues with a higher than average proportion of single tickets. These two orchestras also have some of the lowest scores in Factor 2, suggesting that their audiences respond better than average to Brahms, Turn of Century, and Pre-Romantic, but worse than average to star soloists and popular works. This could mean that the single-ticket audiences for these orchestras are more devoted or knowledgeable about classical music, or that the marketing for “popular” concerts is less aggressive than it would be in other cities. In contrast, DET and PIT also occupy more of their venues (relative to the average orchestra) with single-ticket buyers (positive scores on Factor 1), but they are located further in the vector directions for Prokofiev, Beethoven and Star Soloist (see Fig. 1), suggesting that their audiences respond better than average to these program features.

Further in that direction have larger than average intercepts, and therefore tend to occupy a higher percentage of their venues with single-ticket buyers, all else being the same. These differences in intercepts across orchestras also account for unobservable factors such as season subscriptions that systematically cause single-ticket occupancy differences across markets. Orchestras located in the opposite direction have lower than average single-ticket sales as a proportion of seating capacity. This finding could be due to such reasons as a venue that is too large for the market, a large base of season-ticket holders, or strong

### Table 4
Parameter estimates for the complete model with two latent factors for audience heterogeneity and two dimensions capturing feature interactions.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Average</th>
<th>Audience Heterogeneity Factor 1</th>
<th>Factor 2</th>
<th>Feature Interactions</th>
<th>Interact1</th>
<th>Interact2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.125</td>
<td>0.950</td>
<td>-0.084</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sat eve</td>
<td>0.182</td>
<td>-0.169</td>
<td>0.002</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sun</td>
<td>0.003</td>
<td>-0.116</td>
<td>-0.097</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sun-Thru eve</td>
<td>-0.141</td>
<td>-0.343</td>
<td>-0.191</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jan, Feb, Mar</td>
<td>0.018</td>
<td>0.090</td>
<td>0.027</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Apr, May Jun</td>
<td>-0.053</td>
<td>0.164</td>
<td>0.040</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sept, Oct</td>
<td>-0.146</td>
<td>0.060</td>
<td>-0.111</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Price</td>
<td>-0.014</td>
<td>0.040</td>
<td>0.020</td>
<td>-0.005</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Star soloist</td>
<td>0.142</td>
<td>0.120</td>
<td>0.108</td>
<td>-0.206</td>
<td>-0.298</td>
<td>-</td>
</tr>
<tr>
<td>Freq soloist</td>
<td>-0.081</td>
<td>-0.029</td>
<td>-0.038</td>
<td>-0.715</td>
<td>-0.150</td>
<td>-</td>
</tr>
<tr>
<td>Guest conductor</td>
<td>-0.122</td>
<td>0.023</td>
<td>0.029</td>
<td>-0.039</td>
<td>0.128</td>
<td>-</td>
</tr>
<tr>
<td>Most pop piece</td>
<td>0.242</td>
<td>-0.071</td>
<td>0.123</td>
<td>0.057</td>
<td>0.341</td>
<td>-</td>
</tr>
<tr>
<td>Other pop piece</td>
<td>0.242</td>
<td>-0.067</td>
<td>0.084</td>
<td>0.046</td>
<td>-0.081</td>
<td>-</td>
</tr>
<tr>
<td>Choral</td>
<td>0.226</td>
<td>-0.203</td>
<td>-0.057</td>
<td>0.313</td>
<td>0.359</td>
<td>-</td>
</tr>
<tr>
<td>Mozart</td>
<td>0.083</td>
<td>-0.190</td>
<td>-0.116</td>
<td>-0.055</td>
<td>-0.092</td>
<td>-</td>
</tr>
<tr>
<td>Beethoven</td>
<td>-0.042</td>
<td>0.174</td>
<td>0.099</td>
<td>-0.114</td>
<td>0.037</td>
<td>-</td>
</tr>
<tr>
<td>Tchaikovsky</td>
<td>0.057</td>
<td>-0.113</td>
<td>-0.011</td>
<td>-0.091</td>
<td>-0.035</td>
<td>-</td>
</tr>
<tr>
<td>Brahms</td>
<td>-0.023</td>
<td>-0.301</td>
<td>-0.211</td>
<td>-0.151</td>
<td>0.179</td>
<td>-</td>
</tr>
<tr>
<td>Strauss</td>
<td>-0.190</td>
<td>0.038</td>
<td>-0.041</td>
<td>-0.099</td>
<td>0.154</td>
<td>-</td>
</tr>
<tr>
<td>Prokofiev</td>
<td>-0.088</td>
<td>0.178</td>
<td>0.195</td>
<td>-0.075</td>
<td>-0.077</td>
<td>-</td>
</tr>
<tr>
<td>Pre-Romantic</td>
<td>-0.119</td>
<td>-0.227</td>
<td>-0.208</td>
<td>-0.258</td>
<td>0.048</td>
<td>-</td>
</tr>
<tr>
<td>Romantic</td>
<td>-0.130</td>
<td>-0.025</td>
<td>0.064</td>
<td>-0.118</td>
<td>0.149</td>
<td>-</td>
</tr>
<tr>
<td>Late Romantic</td>
<td>-0.169</td>
<td>-0.087</td>
<td>-0.058</td>
<td>-0.222</td>
<td>0.368</td>
<td>-</td>
</tr>
<tr>
<td>Turn of century</td>
<td>-0.032</td>
<td>-0.192</td>
<td>-0.147</td>
<td>-0.059</td>
<td>0.017</td>
<td>-</td>
</tr>
<tr>
<td>20th century</td>
<td>-0.095</td>
<td>-0.034</td>
<td>-0.085</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>Contemporary</td>
<td>0.008</td>
<td>-0.163</td>
<td>-0.010</td>
<td>0.024</td>
<td>-0.102</td>
<td>-</td>
</tr>
</tbody>
</table>

Values in bold are statistically significant at the p<0.01 level.

### Table 5
Average single-ticket occupancy elasticities for program features.

<table>
<thead>
<tr>
<th>Program feature</th>
<th>% change in occupancy rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturday evening</td>
<td>13.3</td>
</tr>
<tr>
<td>Sunday afternoon</td>
<td>-2.8</td>
</tr>
<tr>
<td>Sunday or Thursday evening</td>
<td>-2.1</td>
</tr>
<tr>
<td>Jan, Feb, Mar</td>
<td>5.0</td>
</tr>
<tr>
<td>Apr, May Jun</td>
<td>0.4</td>
</tr>
<tr>
<td>Sept, Oct</td>
<td>-11.6</td>
</tr>
<tr>
<td>10% cut in quality-adjusted price</td>
<td>0.4</td>
</tr>
<tr>
<td>Star soloist</td>
<td>19.3</td>
</tr>
<tr>
<td>Popular soloist</td>
<td>-6.8</td>
</tr>
<tr>
<td>Guest conductor</td>
<td>-11.0</td>
</tr>
<tr>
<td>Most popular piece</td>
<td>45.7</td>
</tr>
<tr>
<td>Other popular piece</td>
<td>22.6</td>
</tr>
<tr>
<td>Choral</td>
<td>11.0</td>
</tr>
<tr>
<td>Mozart</td>
<td>3.8</td>
</tr>
<tr>
<td>Beethoven</td>
<td>3.5</td>
</tr>
<tr>
<td>Tchaikovsky</td>
<td>3.7</td>
</tr>
<tr>
<td>Brahms</td>
<td>-7.0</td>
</tr>
<tr>
<td>Pre-Romantic</td>
<td>-11.0</td>
</tr>
<tr>
<td>Romantic</td>
<td>-9.3</td>
</tr>
<tr>
<td>Late Romantic</td>
<td>-9.2</td>
</tr>
<tr>
<td>Turn of century</td>
<td>-5.8</td>
</tr>
<tr>
<td>20th century</td>
<td>-9.4</td>
</tr>
<tr>
<td>Contemporary</td>
<td>-7.2</td>
</tr>
<tr>
<td>Strauss</td>
<td>-9.5</td>
</tr>
<tr>
<td>Prokofiev</td>
<td>-2.2</td>
</tr>
</tbody>
</table>
The interactions implied by the interaction map in Fig. 3 are reported in Table 6, where one can see a few statistically significant interactions. For example, Table 6 confirms the negative interaction between Star Soloist and Most Popular piece, suggesting that people who are attracted to a concert by a star soloist may tend to be the same who would be attracted by the most popular pieces. On average both features produce more single-ticket sales, but the inclusion of both features on the same program does not result in an additive effect. The same is true of the features Choral and Star Soloist. It is possible that choristers' friends and family tend to purchase single tickets, but it may be true that they are also attracted to the concert hall when a well-known star soloist is also on the program. Combining these two program features produces an overall effect that is less than the sum of the two individual effects because the audience attracted to the concert hall by Choral overlaps with the audience attracted by Star Soloist.

On the other hand, the positive interaction between Star Soloist and Freq Soloist suggests that featuring a star soloist who has also been frequently featured in past concerts produces a better result. In other words, this positive interaction suggests that for soloists, familiarity leads to audience loyalty. It is also interesting to note that the slightly negative average effects of Romantic music, Late Romantic music, and the neutral average effect of the music of Brahms on a program (see Table 3) show a positive interaction when these repertoires are combined (Fig. 3 and Table 6), suggesting that while each of these items have a neutral or negative average effect, combining them tends to improve the overall attractiveness of the program.

2.0.2. Predictive validity tests

Even though the main goal in developing the proposed model was to obtain a better understanding of audience preferences for concert features, we test for the predictive validity of the proposed model and verify the consequences of over-specification. We apply the parameter estimates obtained from the calibration sample to predict log-occupancy for the last five concerts for each of the 47 orchestras that were withheld from the dataset. Obviously, there are many factors affecting ticket sales for a concert (beyond the program features and scheduling information included in our model), such as competing entertainment options at the same time in the same market, weather conditions at the time of the event, and marketing efforts to promote the season series. Therefore, the predictive performance of this model should be viewed with its main purpose of uncovering preferences for concert features in mind rather than for producing accurate ticket sales forecasts. Despite these caveats, the predictive fit results we discuss next are quite reasonable and strongly support the predictive validity of our proposed framework. Moreover, they show considerable improvement with our proposed framework relative to simpler models that ignore the various aspects captured by our model.

The fit statistics shown in Table 7 confirm that accounting for heterogeneity across audiences has a considerable effect on model performance. Overall, the best predictive performance is obtained with two factors accounting for preference heterogeneity across audiences and one dimension uncovering the interactions among program features. As one would expect, there is some degradation in fit between model calibration and holdout forecasts. However, the
performance in both samples is reasonably close, suggesting stability in the model specification.

2.0.3. Implementation for a non-sampled orchestra

The discussion above focused on the insights the model provides regarding the average appeal of a particular concert program, and on deviations in preferences around this average for the 47 orchestras in our sample. However, the results from our proposed model can also be utilized for an orchestra not included in that sample. Without any prior information regarding how her audience might depart from the average, the marketing director for this orchestra can use the average estimates, reported under “intercept” and “feature interactions” in Table 4, which provide insights into how a “typical” or average audience would respond to the program being considered, as shown in Eq. (4).

On the other hand, if the marketing director believes there are some similarities between her audience and some of the 47 orchestras portrayed in Fig. 3, she could use this information to guess the
likely factor scores for her audience using the scores (map positions) for the audiences in Fig. 3 she believes are similar to her own. This “guestimate” of the scores for her audience would produce estimates of a program’s appeal that are more customized to her particular audience (utilizing Eq. (5)) than the average appeal from Eq. (4). As single-ticket sales data become available for her orchestra, the marketing director can obtain data-driven estimates of the factor scores for her audience following the calculations we describe for the E-step right after Eq. (3).

As an illustration, we consider an anonymous orchestra for which we have data on 15 concerts, not included in our calibration sample. First, we produce forecasts based on the preferences of the typical audience. Then, we attempt to customize the forecasts based on our own subjective (admittedly naive) assessment of the similarities between this orchestra and the ones in our calibration sample. Finally, we use ticket sales data from the first 3, 6 and 9 concerts to obtain factor scores for this new orchestra, which we then use to produce forecasts for the remaining periods. Table 8 and Fig. 4 compare the forecasting performance obtained with the proposed model under these different conditions. From this figure and table one can see that the forecasting performance is substantially improved as sales data become available to compute the audience’s latent scores, which show how this audience departs from the average across all orchestras in the calibration sample. Once the scores for the new orchestra are better estimated, the forecasting performance of the model stabilizes at a much lower prediction error than the a priori forecasts.

3. Conclusions and final discussion

The forecasting model developed and tested in this study illustrates how an organization that produces events by combining interacting features can learn about its audience’s preferences and, in this process, predict ticket sales with greater accuracy. Non-profit performing arts organizations in particular can benefit from our approach by minimizing risk associated with scheduling performers, determining the content and context of these performances, commissioning contemporary pieces, and recruiting guest artists while meeting organizational goals. Following our proposed framework, a similar model could be developed for independent theater companies, dance companies, or opera companies, that would produce insights into their audience’s tastes, while
accounting for interactions among program features and for preference heterogeneity across markets. Using a non-repertoire-specific set of variables, the model could be useful for venues that mix genres and performance types, such as civic centers or community performance spaces. And, with a modified dependent variable, exhibition spaces and museums could employ a version of our model as well. Because it accounts for interactions among program features, our proposed framework allows the organization to anticipate the potential synergies produced by combining certain features or the potential cannibalization of audience resulting from adding a certain feature to a program. For example, a contemporary art museum might want to know the ideal standing these interactions across other contemporary museums would help in this type of decision. Because it allows for heterogeneity in preferences across markets via a factor-regression random-coefficients formulation, our framework allows each organization to “borrow” information from similar organizations, leading to a more robust insight into the organization's own audience and a more reliable assessment of the interactions among program features. A successful performance season at a performing arts center in Dallas might differ dramatically from a successful mix of offerings at a performing arts center in Boston.

Specifically, this study provides insights that could be useful to an orchestra's music and marketing directors about general audience preferences for concert features. Although some of the program features in the study were, as expected, positively or negatively correlated with single-ticket occupancy (see Table 4), a few surprising general conclusions about orchestra single-ticket sales may be drawn from the data:

- the inclusion of a soloist with frequent engagements does not typically result in higher occupancy,
- the inclusion of a guest conductor on the program has a significant negative impact on occupancy,
- less popular works by famous composers such as Beethoven, Brahms, and Tchaikovsky do not necessarily translate into higher attendance,
- less popular works from before the turn-of-the-century have a stronger negative effect on occupancy than less popular works from after the turn-of-the-century, and perhaps most surprisingly,
- contemporary music is the only category of less popular works that does not have a significant negative effect on single-ticket occupancy.

The more predictable results of the model include the positive effects of choral music and popular and famous works and soloists on occupancy, the low occupancy rate that typically occurs at the beginning of a season (September–October), the popularity of Saturday evening, and the lower occupancy that occurs on weekdays.

The interaction effects shown in Fig. 3 are also informative for marketing and music directors. The interaction effects of chief interest are:

- the positive interaction between choral works and the most popular pieces, implying that the joint effect of including a choral work and a “most popular” piece on the same program is greater than the sum of their individual effects;
- the negative interaction between works featuring a star/frequent soloist and popular or choral works, implying that the joint effect of including a star/frequent soloist on the same program as a popular or choral work is less than the sum of their effects; and
- the positive interactions between many of the characteristics of the orchestra’s less popular core repertoire (Pre-Romantic, Romantic, Late Romantic, Brahms, Strauss, Beethoven), suggesting that the combination of these types of works is more effective than the inclusion of a single work of this type on a program.

Our empirical results suggest several programming strategies for music directors who are concerned with maintaining steady single-ticket sales. The interaction effects observed in our model imply that occasional concertgoers are more likely to attend a program that has both a choral work and a very popular piece than a program that has only one or the other. The pairing of works in these categories may result in an overall increase in single-ticket sales. Furthermore, concertgoers who are inclined to attend a program featuring a star (or frequent) soloist are often the same concertgoers who are inclined to attend a program featuring a choral or popular work. Including a star soloist on a program that already includes a choral work or a popular work will not significantly increase single-ticket sales. Finally, combining less popular works (from 1750 to 1950) together on the same concert may (counter-intuitively) be the best way to maximize single-ticket sales for programs of this type. Such a program would, however, tend to sell fewer tickets than a program with a popular work included.

The average effects have additional implications for the business-minded marketing director. For example, single-ticket buyers seem to not be interested in attending performances that feature a guest conductor. It would be wiser, perhaps, to emphasize the resident conductor’s role as a figure in the community and increase sales by marketing him or her as a personality. This approach has achieved success in some markets (Ankeny and Neeme, 2005). Additionally, the inclusion of a featured soloist on a program does not necessarily translate into higher single-

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>Variance</th>
<th>R-square (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>0.44</td>
<td>0.51</td>
<td>14.5</td>
</tr>
<tr>
<td>Aggregate + 1D interactions</td>
<td>0.45</td>
<td>0.51</td>
<td>12.6</td>
</tr>
<tr>
<td>Aggregate + 2D interactions</td>
<td>0.45</td>
<td>0.51</td>
<td>14.1</td>
</tr>
<tr>
<td>1D Random coefficients</td>
<td>0.29</td>
<td>0.51</td>
<td>43.3</td>
</tr>
<tr>
<td>2D Random coefficients</td>
<td>0.29</td>
<td>0.51</td>
<td>44.2</td>
</tr>
<tr>
<td>3D Random coefficients</td>
<td>0.28</td>
<td>0.51</td>
<td>45.3</td>
</tr>
<tr>
<td>2D Random coeff + 1D interactions</td>
<td>0.27</td>
<td>0.51</td>
<td>46.7</td>
</tr>
<tr>
<td>2D Random coeff + 2D interactions</td>
<td>0.29</td>
<td>0.51</td>
<td>44.4</td>
</tr>
</tbody>
</table>

Note: MSE = mean squared error.

The interaction effects shown in Fig. 3 are also informative for marketing and music directors. The interaction effects of chief interest are:

- the positive interaction between choral works and the most popular pieces, implying that the joint effect of including a choral work and a “most popular” piece on the same program is greater than the sum of their individual effects;
- the negative interaction between works featuring a star/frequent soloist and popular or choral works, implying that the joint effect of including a star/frequent soloist on the same program as a popular or choral work is less than the sum of their effects; and
- the positive interactions between many of the characteristics of the orchestra’s less popular core repertoire (Pre-Romantic, Romantic, Late Romantic, Brahms, Strauss, Beethoven), suggesting that the combination of these types of works is more effective than the inclusion of a single work of this type on a program.

Our empirical results suggest several programming strategies for music directors who are concerned with maintaining steady single-ticket sales. The interaction effects observed in our model imply that occasional concertgoers are more likely to attend a program that has both a choral work and a very popular piece than a program that has only one or the other. The pairing of works in these categories may result in an overall increase in single-ticket sales. Furthermore, concertgoers who are inclined to attend a program featuring a star (or frequent) soloist are often the same concertgoers who are inclined to attend a program featuring a choral or popular work. Including a star soloist on a program that already includes a choral work or a popular work will not significantly increase single-ticket sales. Finally, combining less popular works (from 1750 to 1950) together on the same concert may (counter-intuitively) be the best way to maximize single-ticket sales for programs of this type. Such a program would, however, tend to sell fewer tickets than a program with a popular work included.

The average effects have additional implications for the business-minded marketing director. For example, single-ticket buyers seem to not be interested in attending performances that feature a guest conductor. It would be wiser, perhaps, to emphasize the resident conductor’s role as a figure in the community and increase sales by marketing him or her as a personality. This approach has achieved success in some markets (Ankeny and Neeme, 2005). Additionally, the inclusion of a featured soloist on a program does not necessarily translate into higher single-

The interaction effects shown in Fig. 3 are also informative for marketing and music directors. The interaction effects of chief interest are:

- the positive interaction between choral works and the most popular pieces, implying that the joint effect of including a choral work and a “most popular” piece on the same program is greater than the sum of their individual effects;
- the negative interaction between works featuring a star/frequent soloist and popular or choral works, implying that the joint effect of including a star/frequent soloist on the same program as a popular or choral work is less than the sum of their effects; and
- the positive interactions between many of the characteristics of the orchestra’s less popular core repertoire (Pre-Romantic, Romantic, Late Romantic, Brahms, Strauss, Beethoven), suggesting that the combination of these types of works is more effective than the inclusion of a single work of this type on a program.

Our empirical results suggest several programming strategies for music directors who are concerned with maintaining steady single-ticket sales. The interaction effects observed in our model imply that occasional concertgoers are more likely to attend a program that has both a choral work and a very popular piece than a program that has only one or the other. The pairing of works in these categories may result in an overall increase in single-ticket sales. Furthermore, concertgoers who are inclined to attend a program featuring a star (or frequent) soloist are often the same concertgoers who are inclined to attend a program featuring a choral or popular work. Including a star soloist on a program that already includes a choral work or a popular work will not significantly increase single-ticket sales. Finally, combining less popular works (from 1750 to 1950) together on the same concert may (counter-intuitively) be the best way to maximize single-ticket sales for programs of this type. Such a program would, however, tend to sell fewer tickets than a program with a popular work included.

The average effects have additional implications for the business-minded marketing director. For example, single-ticket buyers seem to not be interested in attending performances that feature a guest conductor. It would be wiser, perhaps, to emphasize the resident conductor’s role as a figure in the community and increase sales by marketing him or her as a personality. This approach has achieved success in some markets (Ankeny and Neeme, 2005). Additionally, the inclusion of a featured soloist on a program does not necessarily translate into higher single-
ticket sales, even if that soloist has numerous engagements with other orchestras. Only star soloists (i.e., chart-topping recording artists) are likely to draw additional single-ticket buyers. Finally, single-ticket buyers seem to be less interested in works from before 1900 – even those considered to be standard repertoire – unless the work can be considered a “popular” piece. Related to this phenomenon is the non-negative effect of contemporary compositions on single-ticket sales. If marketed in the right way, contemporary music has the potential to draw larger audiences of single-ticket buyers than Romantic Era music. Just as single-ticket buyers seem to prefer and identify with their local conductor, they may more easily identify with composers of today (especially local composers) than with composers of two hundred years ago.

It is clear from Fig. 2, however, that significant variation in preference exists across markets. This figure, combined with Fig. 1 provides direct insights into how each audience differs from the sample average, helping the music director to tailor a concert program to the specific tastes of her audience. Most importantly, as we demonstrated in our empirical illustration, managers from any orchestra included in our sample can obtain parameter estimates that reflect the tastes of their specific audience. Additionally, as demonstrated in the empirical illustration, managers from orchestras not included in our sample may utilize the model by locating their audience in Fig. 2 based on their similarity to other audiences or by gathering historical data on concert features and ticket sales for their own orchestra.

In this study, we focused on single-ticket sales because they are more likely to depend on the features of each concert, as evidenced by the fact that most participating orchestras were unable to provide subscription data at the concert level. It should be noted that this is a limitation of the data available to us rather than of the proposed model. In fact, our random-coefficients formulation attempts to minimize the biasing effect of these unobservable season tickets by incorporating a random component capturing unobservable orchestra-level effects. Future applications could extend our proposed forecasting framework to include season ticket sales, which would have to be broken down to the event level so they can be directly related to program features.

Because we did not have data on subscription sales for each event, we also ignored the possibility that some events might have been sold-out of single tickets, something that has usually been overlooked in the performing-arts literature, with the notable exception of Putler and Lele (2003). Our sources in the industry indicate that sold-out regular-season concerts are uncommon. However, as data on all ticket sales become available for each event, a truncated extension of our proposed framework will be necessary to account for this truncation in ticket sales due to venue capacity constraints.

As noted previously, our effort focused on how audiences respond to program features, assuming these features as exogenous variables and ignoring the possibility that music/program directors have already developed an intuition about their audiences’ preferences and behaviors, using this intuition to design and schedule programs to maximize utilization of their venues. Therefore, program features could be endogenous to a supply–demand system, rather than exogenous as we assumed in our demand model. The real problem is much more complex than what one would expect in theory because music/program directors must operate under many external constraints. Of course, to engage a highly coveted soloist or guest conductor, scheduling adjustments may be necessary. More importantly, most orchestras do not own their own venues, and they must first construct the season’s schedule subject to the availability of the hall. Additionally, the music director must be available for most or all of the concerts. If the orchestra has a Pops series, the regular season Classical subscription series must be scheduled so as not to conflict with the Pops series. The availability of the Pops conductor and performers would then need to be taken into account as well.

The availability of the orchestra’s players is also crucial, and this varies from orchestra to orchestra because larger orchestras have salaried players who are expected to attend all rehearsals and performances, whereas smaller and mid-size orchestras have players under contract for a certain number of weeks or services. This means that for most orchestras, the addition of an extra service in a given week can be prohibitively expensive, or requires the elimination of a service elsewhere in the season’s schedule. Furthermore, in many smaller and mid-size markets orchestral players perform and rehearse with other orchestras in nearby cities (e.g., Portland, ME and Providence, RI), so schedules are sometimes constructed to allow the musicians additional performance opportunities. In Portland, for example, there is a long tradition of Tuesday night concerts because that is when players happen to be available.

Unfortunately, we could not find any historical register of the context in which programming and scheduling decisions were made by the 47 orchestra directors (with as many as 619 concerts each). A retrospective re-construction of this record would be impractical and fraught with errors and inconsistencies. Therefore, modeling programming/scheduling decisions and audience response simultaneously was not feasible with the available data. Nevertheless, a more comprehensive

![Fig. 4. Predicted single-ticket sales (as a % of capacity) for a new orchestra based on prior and updated estimates.](image-url)
equilibrium model that would simultaneously account for the demand and supply decisions made by all audiences and all planners would be a logical next step, as more comprehensive data are gathered from the supply side. We leave this as an opportunity for future research.

Finally, although our focus in this study was on the performing arts and symphony orchestras in particular, the main features of our proposed framework can be potentially useful in many other marketing contexts in which managers must assemble multiple, potentially interacting attributes to attract customers who may respond differently to these attributes in different markets. Our framework allows the manager to uncover the diversity in “tastes” for product/service features across markets while accounting for potential synergy and cannibalization among these features. We hope the framework will also prove useful and insightful for marketers offering different bundles of features in multiple markets (e.g., fast-food chains, grocery retailers and hotel chains).

References


The effects of mailing design characteristics on direct mail campaign performance

Sebastian Feld a, Heiko Frenzen b, Manfred Krafft c, Kay Peters d,⁎, Peter C. Verhoef e,⁎

a Center for Interactive Marketing and Media Management, c/o Institute of Marketing, University of Muenster, Am Stadtgraben 13-15, 48143 Muenster, Germany
b Aston Business School, Aston Triangle, Birmingham B4 7ET, United Kingdom
c Institute of Marketing, University of Muenster, Am Stadtgraben 13-15, 48143 Muenster, Germany
d SVI-Endowed Chair for Dialog Marketing, Institute of Marketing and Media, University of Hamburg, Esplanade 36, 20354 Hamburg, Germany
e Faculty of Economics and Business, University of Groningen, 9747 AE Groningen, Netherlands

A R T I C L E   I N F O

Article history:
First received in 18, June 2011 and was under review for 6 months
Available online 18 October 2012
Area Editor: Harald J. van Heerde

A B S T R A C T

Designing effective direct mail pieces is considered a key success factor in direct marketing. However, related published empirical research is scarce while design recommendations are manifold and often conflicting. Compared with prior work, our study aims to provide more elaborate and empirically validated findings for the effects of direct mail design characteristics by analyzing 677 direct mail campaigns from non-profit organizations and financial service providers. We investigate the effects of (1) various envelope characteristics and observable cues on opening rates, and (2) characteristics of the envelope content on the keeping rates of direct mail campaigns. We show that visual design elements on the outer envelope – rather than sender-related details – are the predominant drivers of opening rates. Factors such as letter length, provision of sender information in the letter, and personalization positively influence the keeping rate. We also observe that opening and keeping rates are uncorrelated at the campaign level, implying that opening direct mail pieces is only a necessary condition for responding to offers, but not per se a driver of direct mail response.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Direct marketing is a key component of the advertising media mix for many firms (DMA, 2011). Direct marketing serves a range of firm communication goals from creating brand awareness to generating response along with TV, print or online advertising (e.g., Briggs, Krishnan, & Borin, 2005; Naik & Peters, 2009). Among all direct marketing media, direct mail is clearly the predominant element, accounting for over one-third of direct marketing expenditures in most countries (DMA, 2011). Hence, consumers are confronted with a continuously growing direct mail volume in the mailbox that brings increased competition for their limited attention (Van Diepen, Donkers, & Franses, 2009a). In responding to this competition, firms follow two primary routes. First, they improve the targeting, timing, and sequencing of their direct mail campaigns. This development is well reflected in the academic literature, which explores how the response to direct mail has been optimized by better segmentation and targeting (e.g., Bult & Wansbeek, 1995; Donkers, Paap, Jonker, & Franses, 2006) as well as better timing and sequencing, and by identifying the appropriate number of mailings per customer (e.g., Elsner, Krafft, & Huchzermeier, 2004; Cönül & Ter Hofstede, 2006; Jen, Chou, & Allenby, 2009; Rust & Verhoef, 2005; Van Diepen et al., 2009a).

Second, to catch attention in the mailbox, firms strive to improve the design of their direct mail. These efforts are reflected in the increased focus that design characteristics are given in practice and in the textbooks on direct marketing (e.g., Nash, 2000; Stone & Jacobs, 2008). The textbooks claim that the creative elements and design characteristics of direct mail account for up to a quarter of its overall success (e.g., Roberts & Berger, 1999, p. 7; Stone & Jacobs, 2008, p. 6): the favorable presentation of the solicitation facilitates the consumer's response process by attracting attention and then generating interest in the offer. Accordingly, design primarily acts as a critical response enabler in the early and intermediate stages of the direct mail funnel. Hence, design drives the intermediate stages, such as the opening and reading of a direct mail piece, rather than ultimate response (De Wulf, Hoekstra, & Commandeur, 2000). Unfortunately, these pre-response stages are currently a “black box” for marketing managers: they only observe the number of final responses resulting from a particular campaign. This limitation could, to some extent, explain why design optimization has received less attention compared to selection and targeting, both of which are easily measured and can be linked directly to response. If systematic marketing research data on the intermediate funnel stages were available, however, such information could provide diagnostic value to managers. Given the low response rates of approximately 1–2% on average (DMA, 2006), it would help to infer where and why the majority of direct mail becomes stuck in the direct mail funnel and how to overcome it. This study analyzes a unique...
commercial direct mail panel that explicitly covers these intermediate direct mail funnel stages by measuring the opening and keeping rate.

In general, there are numerous specific mailing design guidelines in the practitioner literature without an emerging consensus (e.g., about whether cover letters should be short or long) and mostly without reference to any empirical study. Only a few scientific empirical studies have been published on related issues. These studies focus mostly on the advertising context of direct marketing and on a particular industry or firm, and they typically employ firm-specific experimental designs leading to non-generalizable results (e.g., Bell, Ledolter, & Swersey, 1997; De Wulf, 2000; Seaver & Simpson, 1995; Vriens et al., 1996). Given the variety of industries and their peculiarities (e.g., Stone & Jacobs, 2008), it would not be surprising if only a few design characteristics actually achieve cross-industry importance, while the majority of effects might be industry-specific.

Accordingly, our primary research objective is to investigate the effect of direct mail design on the intermediate stages of the direct mail response funnel, namely the opening and keeping rates of direct mailing campaigns instead of the ultimate response rates. As a second research objective, we intend to compare the effect of specific design characteristics across industries, investigating the extent to which findings in one industry can be valid for another. Accordingly, we use a database of 677 direct mail campaigns in 2 industries: the financial and the non-profit industry. For the intermediate stages, we define the opening rate (OR) of a campaign as the percentage of recipients that open the direct mail envelope, while the keeping rate (KR) is defined as the percentage of recipients that keep the mailing after opening the envelope. Using these intermediate communication metrics at the campaign level will enable us to shed unprecedented light into the “black box” of the direct mail funnel.

Our empirical results show that the design elements substantially impact the OR and the KR. Surprisingly, we observe no relationship between the OR and the KR, implying that opening is a necessary but not sufficient condition for generating a campaign response. Our results also show that some design characteristics are of varying importance at different stages of the direct mail funnel. For instance, presenting the sender’s logo on the envelope decreases the OR for financial service providers. Providing this information in the letter, however, increases the KR in both industries.

The remainder of the paper is organized as follows. We review the literature on direct mail design. Next, we present our research framework and its theoretical underpinnings. After this, the data collection, sample properties, and model estimation will be described, followed by a presentation of our empirical results. From these results, we will derive conclusions as well as implications for research and management. We conclude with directions for further research.

2. Prior research on direct mail design

We define the scope of our literature review based on 2 selection criteria. First, the studies should focus on the effects of the design characteristics rather than on the other success factors of the direct mail solicitations (such as timing, targeting, offer design, or message appeals). Second, we exclude studies on the design of mail surveys (e.g., Gendall, 2005; Helgeson, Voss, & Terpening, 2002; Yu & Cooper, 1983). Hence, we select only studies that involve design features in commercial direct mail solicitations. Table 1 provides an overview of the studies published in reviewed journals that fit these criteria. Across the studies, we compare (1) the research design and sample description, (2) the dependent variables, and (3) the category of independent variables, namely design characteristics and covariates. At this point, we intentionally refrain from describing the empirical results of these studies. Instead, we will draw on their findings later in Section 4.2 when discussing the effects of different types of design characteristics. Here, we focus on the methodological aspects to highlight the gaps in the previous research that our study aims to address.

2.1. Research design and sample description

The reported studies have typically adopted field experiments with a single firm in a single industry. In particular, non-profit organizations have been frequently studied; the other industries studied are primarily financial or B2B services. The number of investigated campaigns or
different stimuli employed varies between 2 and 20. Given the limited number of campaigns and stimuli investigated in the previous studies, inferring general insights is hardly feasible. Hence, there is a need for a study that considers a large number of campaigns to develop generalizable findings.

2.2. Dependent variables

The direct mail response rate is the most frequently studied behavioral variable (in 14 out of 18 cases). However, the studies investigating envelope design characteristics usually investigate their impact on the OR (e.g., James & Li, 1993; Vriens, van der Scheer, Hoekstra, & Bult, 1998). When the additional design characteristics from the other direct mail elements are included, they are related to reading behavior as an intermediate measure that reflects elevated interest (De Wulf et al., 2000). Additionally, the levels or the variants of the design element varies between 1 (e.g., Capon & Farley, 1976) and 14 (De Wulf et al., 2000). The number of characteristics studied per mail element varies between 2 and 20. Given the limited volume). Not controlling for these effects can potentially cause biased estimates of the design characteristic’s impact.

In sum, the literature review suggests the need for a study that covers a substantial portion of all campaigns rather than only selected mailings from single organizations. The study needs to span the direct mail response funnel from (a) opening behavior to a stage of (b) interest to (c) ultimate response. Design exerts its primary influence on the first 2 stages of the direct mail funnel, while the final stage of actual response is largely driven by targeting, timing, and the actual offer characteristics (De Wulf et al., 2000). Accordingly, our investigation on the design characteristics focuses on the intermediate opening and interest stages of the funnel.

2.3. Independent variables

The design characteristics are usually attributed to 4 core mail elements: (1) the envelope, (2) the cover letter, (3) any supplements (e.g., leaflets, brochures or catalogs), and (4) the response device. Selectively, 2 additional categories comprise add-ons (e.g., enclosure of incentive) and covariates (e.g., characteristics of recipients). These elements contribute differently across the stages of the direct mail funnel. For example, envelope design characteristics and observable haptic cues are the main drivers of the opening behavior because the other elements are usually invisible to the recipients. Hence, we incorporate this distinction into our framework.

The majority of studies investigate a limited number of characteristics across selected mail elements with some notable exceptions (e.g., Bell et al., 2006; Bult, van der Scheer, & Wansbeek, 1997; De Wulf et al., 2000). The number of characteristics studied per mail element varies between 1 (e.g., Capon & Farley, 1976) and 14 (De Wulf et al., 2000). Additionally, the levels or the variants of the design characteristics investigated appear to be largely driven by the specific context of the cooperating organization. To avoid a bias in effect inference, we need to account for a rather comprehensive set of design elements across all 4 core mail elements and derive the levels of the design characteristics from a broader set of sources, e.g., from exploring our panel data, from the literature, and from industry expert interviews.

Only a few studies control for the effects of covariates (e.g., campaign volume). Not controlling for these effects can potentially cause biased estimates of the design characteristic’s impact.

In sum, the literature review suggests the need for a study that covers a substantial portion of all campaigns rather than only selected mailings from single organizations. The study needs to span the direct mail response funnel with a focus on the “black box” that occurs prior to response, and it should employ a comprehensive set of design characteristics extracted from various sources. There is an additional need to control for various covariates that might exert significant influence on the direct mail funnel stages.

3. Conceptual framework

From the literature review, we infer that different funnel stages exist along the direct mail response process. These linear stages can be explicitly linked because the outcome at a specific stage depends on the outcome of the previous. For instance, at the first stage, a
certain percentage of recipients pay elevated attention to the piece and might decide to open the envelope; others discard the mail piece without further attention. The ratio of opening to total recipients can then be defined as the OR. At the second stage, the mail recipients exhibit some level of interest in the other mail elements and read them. The recipients might eventually decide to keep the mail for further action (e.g., response). The percentage of recipients who keep the direct mail piece in relation to the number of recipients who open it can be defined as the KR. This measure thus reflects consumer interest. At the third stage, after deciding to keep the mail piece, the recipients might finally decide to respond to it. This results in the qualified response rate. As the direct mail recipients follow this staged process, they build up their commitment while moving step-by-step towards the offer, inducing them to behave consistently with the small prior commitments they have made. This link at the individual recipient level should be reflected in the linked subsequent stages at the campaign level. Taken together, the overall response rate that the managers usually observe builds up as described in Eq. (1):

Following the extant literature on direct mail effectiveness and information processing, the design characteristics exhibit the strongest effect at the first 2 stages of the direct mail funnel: OR and KR (Broadbent, 1958; De Wulf et al., 2000; Pieters & Wedel, 2004). Additionally, both ratios are necessary predecessors of response and thus provide managers with valuable diagnostics—allegorical to the attention and intention measures used for other media. Similar to TV and print advertisements, direct mail pieces are exposures to stimuli that generate contact with the recipients of these campaigns. Opening a mail item is equivalent to a qualified contact because the envelope and its design create a certain degree of curiosity and interest in further investigating the content of the mail item. Taking a closer look at the letter, the brochure and/or response device at the second stage reflects a larger extent of processing information. This elevated interest towards the offer, expressed in our KR measure, enables the repetition of the sender’s messages, facilitating processing and increasing encoding opportunities. Reading and keeping a piece of mail can nurture the sharing of a firm’s message with others and can help to form brand attitude (MacInnis & Jaworski, 1989). Hence, this study fills an important gap in the direct marketing research, where intermediate communication measures have been studied to only a limited extent (exceptions are De Wulf et al., 2000; Diamond & Iyer, 2007; Vriens et al., 1998).

Our conceptual model is shown in Fig. 1. The model includes an explicit link between OR and KR, as implied by the direct mail funnel specified in Eq. (1). In our model, both dependent variables are driven by design characteristics. We categorize these design characteristics along 2 dimensions. First, we group the general design characteristics according to the mail element as performed in previous studies: (1) envelope, (2) letter, (3) supplement, and (4) response device. We assume that the envelope characteristics exert a direct influence on the OR by definition. It is conceivable, however, that the direct mail recipients are able to gain a sense of the contents of the mail package even before opening the envelope, resulting in observational learning. To capture these haptic experiences, we incorporate several envelope content features when analyzing the OR, such as weight, supplements, or give-aways that might be sensed before opening. The design characteristics of the other mail elements cannot be observed at that time. Accordingly, the design characteristics of the other mail elements are assumed to influence the KR. Second, within each mail element, we categorize each design characteristic by its dominating nature, i.e., whether it (1) constitutes a visual design element, (2) identifies the originating sender, (3) represents a personalization cue, or (4) is a measure of information intensity. Third, as suggested in the literature, we investigate the effects of additional industry-specific characteristics along the mail elements, e.g., the position of the payment device in the case of charitable

---

**Table 1 (continued)**

<table>
<thead>
<tr>
<th>Study</th>
<th>Dependent variable(s)</th>
<th>Design characteristics and covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1: Opening</td>
<td>Stage 2: Interest</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Envelope</td>
<td>Letter</td>
</tr>
<tr>
<td>Capon and Farley (1976)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hozer and Robles (1985)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Beard et al. (1990)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gordon and Kellerman (1990)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Sherman, Greene, and Plank (1991)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Williams, Beard, and Kelly (1991)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>James and Li (1993)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>McMellon and Ducoffe (1995)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Seaver and Simpson (1995)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Smith and Berger (1996)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Van der Scheer et al. (1996)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bult et al. (1997)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Vriens et al. (1998)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>De Wulf et al. (2000)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bell et al. (2006)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ledolter and Sweesey (2006)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bekkers and Crutzen (2007)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diamond and Iyer (2007)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Current study (2013)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 1: Overview of studies focusing on direct mail effectiveness and information processing. The design characteristics and covariates are categorized into: (1) envelope, (2) letter, (3) supplement, and (4) response device. The dependent variables are focused on the response rate. ---
mailings from Non-profit Organizations (NPOs) or information on the nearest branch for a financial service provider (FSP). Extending the mail element categories above, we add information on the included incentives (NPO) and offer-related information in both industries. These additional categories add contextual information that could either moderate the effect of the design characteristics or could have a direct effect on our dependent variables. Fourth, we introduce 3 covariates to control for the main drivers of mail performance apart from design: (1) the relative campaign volume within the respective industry, reflecting the selection approach; (2) the share of voice or annual sender volume, reflecting the sender’s position in the respective industry and advertising channel; and (3) the “end-of-month” effect, as keeping behavior is most likely higher for the direct mail received at the end of the month because consumers have had less time to respond before sending it in for collection purposes and therefore tend to keep it. In the following section, we elaborate on the underlying theory and the effects of the direct mail design characteristics on the OR and the KR.

4. Theory on the effects of the design characteristics

4.1. Theoretical foundation

The inclusion of the 4 types of design characteristics discussed above can be motivated by the capacity theories of attention (e.g., Broadbent, 1958) as well as by information-processing models (e.g., MacInnis & Jaworski, 1989). According to Broadbent’s (1958) filter theory, a consumer’s perceptual system contains a filter mechanism. Among the many stimuli or messages presented, only those stimuli that possess salient physical characteristics are allowed through the filter and are subsequently actively processed. Hence, Broadbent’s theory helps to explain the selectivity of attention. The theory implies that salient and familiar verbal or visual stimuli should be used to attract the consumers’ attention. Interestingly, similar insights regarding saliency-based attention have been delivered in the field of neuroscience (e.g., Itti & Koch, 2001) and by related research in marketing (e.g., Van der Lans, Pieters, & Wedel, 2008; Zhang, Wedel, & Pieters, 2009). Examples of salient stimuli in a direct mailing are the use of teasers and headlines, postscripts, typographic accentuations, special envelope formats, colored illustrations or paper, and so forth.

With regard to the information processing models, the processing of an ad stimulus is a function of motivation, ability and opportunity (M-A-O), which are, in part, influenced by the physical properties and design characteristics of the advertisement. More specifically, it is considered that advertisement design properties such as format and size, color, headlines, typography, and other creative elements play a crucial role in attracting consumers’ attention (e.g., Pieters & Wedel, 2004; Pieters, Wedel, & Zhang, 2007) as well as in building persuasive and emotional effects (e.g., Percy & Rossiter, 1983; Smith, Mackenzie, Yang, Buchholz, & Darley, 2007; Yang & Smith, 2009). Based on these theories and research on the effects of design in other media, we assume that the 4 different types of direct mail design categories exhibit differential effects on the direct mail performance, particularly at the first 2 stages of the direct mail funnel.

4.2. The effects of direct mail design categories on the opening and keeping rates

4.2.1. Visual design

Our first category of variables refers to visual design elements such as color, illustrations, bold type or capital letters, extraordinary mailing formats, etc. The use of diverse visual stimuli and their effects on consumers’ reactions has been extensively investigated in the context of print advertisements (e.g., Assael, Kofron, & Burgi, 1967; Percy & Rossiter, 1983; Pieters & Wedel, 2004). In particular, the effects of visual stimuli have been the subject of research in visual imagery (e.g., Rossiter, 1982; Rossiter & Percy, 1980). It has been shown that pictorial stimuli can facilitate persuasive communication in a variety of ways. For example, pictures can lead to more extensive mental processing because they are attention-getting devices (Finn, 1988; Mackenzie, 1986). In addition, pictures can improve the memorability of other semantic information. Research has generally supported the view that pictures can affect ad and brand attitudes, beyond the effects they have on the consumers’ beliefs about the product (e.g., Miniard, Bhatia, Lord, Dickson, & Umnava, 1991). The direct marketing literature has provided some initial support for the effectiveness of using certain visual stimuli such as typographic accentuations and illustrations (Bult et al., 1997), teasers (Roberts & Berger, 1999; Van der Scheer et al., 1996; Vriens et al., 1998), or special envelope formats (Nash, 2000; Vriens et al., 1998).

4.2.2. Sender identity

The second category refers to the presentation of the originating sender’s name and/or logo on the direct mail piece, which translates into the prominence of the brand element. These sender-related cues can be featured on all elements of the direct mail package. Contradicting theories on the effects of sender-related cues can be found in the literature (Pieters & Wedel, 2004). Some scholars argue that a prominent
brand element drives more attention to the brand, which is a necessary condition for obtaining the desired brand-communication effects (e.g., Keller, 2007). In contrast, some advertising practitioners caution against highlighting the brand in advertising because the brand element might signal that the message is an advertisement in which consumers purportedly are not interested (e.g., Aitchinson, 1999; Kover, 1995).

In the context of direct mail design, most of the brand-related debate centers on whether the sender should be clearly displayed on the outer envelope or not. Featuring the sender’s name can signal familiarity and trustworthiness to the recipient (Hoyer & MacInnis, 2007). Conversely, not placing the sender’s name or logo is likely to create curiosity with the direct mail receiver (Nash, 2000; Roberts & Berger, 1999) and might result in higher ORs. However, this immediate effect might be counter-productive if the receiver feels deceived or irritated by the unexpected commercial content of the letter (Nash, 2000; Van Diepen, Donkers, & Fransen, 2009b). The preceding discussion suggests that featuring the sender’s name or brand in promotional campaigns can be dysfunctional. The prior studies on direct mail design did not show any significant effects from revealing the sender’s identity on the envelope on the opening or the response behavior (Bell et al., 2006; De Wulf et al., 2000; Vriens et al., 1998).

4.2.3. Personalization

The third category reflects the degree of personalization for the direct mail design. Personalization is intimately connected with the idea of interactive marketing. Dillman (2007) offers personalization guidelines for surveys that are applicable to direct mail design as well. His personalization strategy is based on the guiding principle that the tone and content of a cover letter should reflect the style used in a business letter to an acquaintance who is not known to the sender. The specific elements of personalization proposed by Dillman are as follows: specific date (e.g., March 14th, 2013); the recipient’s name and address; a personal salutation; a real signature in contrasted ink (i.e., a “pressed blue ball-point pen signature”); and letterhead rather than copied stationery (Dillman, 2007).

Prior research suggests that personalized advertising approaches might increase attention and response to offers (e.g., Ansari & Mela, 2003). However, personalization or customization is not beneficial under all circumstances (e.g., Kramer, Spolter-Weisfeld, & Thakkar, 2007; Zhang & Wedel, 2009) and can even be harmful if the personalized solicitations are perceived as intrusive (e.g., White, Zahay, Thorbjørnsen, & Shavitt, 2008). With regard to direct mail advertising, research has provided moderate support for the positive effects of personalization on response behavior (e.g., Bell et al., 2006; De Wulf et al., 2000; Hozier & Robles, 1985; James & Li, 1993). Studies investigating response rates to mail surveys have yielded mixed findings on personalization: in their review of 93 journal articles, Yu and Cooper (1983) find significant results showing the response-enhancing effects of personalization. In contrast, the more recent survey response studies failed to detect any significant effects from personalization on attention (Helgeson et al., 2002) and response rates (e.g., Gendall, 2005).

4.2.4. Information intensity

The last category of variables refers to the amount of information present in an advertisement. Within the M-A-O-framework, information intensity can affect the recipient’s opportunity to process a message (MacInnis & Jaworski, 1989). From a memory perspective, it seems that the limited capacity of short-term memory is of less concern if the receiver deliberately seeks exposure to the advertisement and actively attends to the content (Rossiter, 1982). For example, direct mail advertising often receives active attention from consumers once the envelope has been opened. Hence, unlike in print or TV ads, the typical technique in direct mail advertising is to provide the reader with sufficient information to achieve a decision to advance the response process (Rossiter, 1982, p. 103). Only Beard, Williams, and Kelly (1990) investigate the effects of information intensity empirically, i.e., response rates of long versus short cover letters in direct mailings, but they find no significant impact.

5. Methodology

5.1. Data and sample description

Our unique data set is based on a representative direct mail panel from GfK. This household panel consists of 3,000 households whose socio-demographics are representative of the entire population of the 35 million private German households. The panel is solely aimed at measuring the intermediate effects, namely the OR and KR, along the direct mail funnel. The panel does not measure the actual response.

The panel participants continuously collect any unsolicited and personally addressed direct mail piece that they receive. At the end of each month, the panel members send GfK all of the direct mailings that they have received during that month and that they do not want to keep. These mailings are either (i) unopened mailings that would normally be discarded right away, or (ii) opened mailings that would be discarded due to a lack of appeal for the recipient after checking the content. GfK scans all of these mail pieces, stores the images in a picture database, and records some key characteristics such as weight, envelope format, postage, or type of response device.

For those mailings that the panel members choose to keep for further consideration (e.g., to read the letter/brochure in greater detail or to respond to the offer at a later time), they are asked to fill out and send GfK a form listing all of these mailings line by line. The specific instruction for the panel participants is as follows: “Below, please fill in only those personally addressed direct mailings that you do not want to send to us, because you want to keep them. Please do not fill in any direct mailings that you send us.” For each of these mail pieces, the panel participant is required to fill in the sender’s name, the date the mailing was received, the type of mailing (postcard, letter or catalog), and the essential subject (slogan/theme) of the campaign. GfK uses this information to precisely match the individual mailings received by households with specific campaigns. The KR of a campaign is then calculated as the percentage of recipients in the panel who keep the corresponding mail piece in relation to the total number of recipients who opened the direct mail piece (see Eq. (1)). For this study, GfK provided us with the aggregate ORs and KRs per campaign derived from this panel as well as access to sample copies of the respective direct mail pieces. The actual response rates are not available because this would require the cooperation of all of the organizations that sent direct mail pieces. These organizations generally regard their actual response rates as very sensitive information.

5.1.1. Sample description

Our sample comprises information on the largest direct mail campaigns (in terms of mailing volume) across a 1-year period from 2 different industries—non-profit organizations (NPOs) and financial service providers (FSPs). Both industries are characterized by a heavy reliance on direct mail campaigns (DMA, 2011; Van Diepen et al., 2009a,b). Together, they account for over 30% of the total mailing volume represented in the GfK direct mail panel and, thus, both belong in the top 5 industries employing direct mailings. Across both industries, we observe 677 distinct campaigns: 396 campaigns (58.5%) from 98 different organizations in the NPO subsample and 281 campaigns (41.5%) from 48 firms in the FSP subsample. Only 1 or 2 campaigns were executed by
54.1% (60.4%) of the NPOs (FSPs); 34.7% (18.8%) of the NPOs (FSPs) ran between 3 and 9 campaigns; and 14.2% (20.8%) of the firms ran 10 or more campaigns within the 1-year period. Of the mailing packages, 97.9% (98.0%), or nearly all, include a cover letter, 71.4% (65.8%) contain a supplement and 97.5% (80.1%) contain a response device (including a payment device in the NPO sample).

5.1.2. Dependent variables—direct mail funnel

GfK records the receipt of the distinct direct mail pieces by household and how many of these pieces of mail were opened or kept. Thus, GfK computes the OR and KR per campaign as described in Eq. (1). The mean OR for the NPO (FSP) campaigns is 87.4% (88.9%). These values are consistent with the evidence from the direct mail literature and practice, indicating the high propensity of consumers to open and read direct mailings (e.g., Nielsen, 2009; Deutsche Post, 2006; Stone & Jacobs, 2008, p. 412). The average KR for the NPO (FSP) industry is 8.2% (5.3%) per campaign. These percentages are very close to the response intention percentages of 8.1% (4.3%) in the NPO (FSP) industry found in a recent U.S. study (DMA, 2011, p. 28), again lending international validity to the German data. The actual response rates are usually substantially lower (1.38–3.42%; DMA, 2011) and sufficiently distinct, thus underlining the importance of the KR as an intermediate measure of the direct mail response funnel.

5.1.3. Independent variables—design elements

To identify and operationalize the design elements for our analysis, we follow a 4-step procedure. First, we search for elements that relate to the theories of attention capacity and information-processing models as well as to our framework by mail element (e.g., envelope) and design characteristic (e.g., personalization). Second, we check the literature in Table 1 and the prominent textbooks (e.g., Geller, 2002; Jones, 1997; Nash, 2000; Roberts & Berger, 1999; Stone & Jacobs, 2008) for cues on the relevant design elements. Third, we scan our database across both industries and collect a variety of design elements empirically. Fourth, we conduct a series of interviews with industry experts that have NPO and FSP backgrounds as well as with specialized advertising agencies, Germany’s largest lettershops, and Deutsche Post DHL. Similarly, our choice for the specific attribute levels is informed. As a result, we arrive at a collection of design variables, their operationalization, and their expected impact on the OR and the KR as shown in Table 2.

Only a few design characteristics (e.g., product category, format, postage and weight) are tracked and recorded by GfK in a systematic fashion. We manually classify and code all other (design) characteristics for each of the 677 campaigns based on the original direct mail piece provided by GfK. The vast majority of design characteristics are rather objective in nature (e.g., presence of teaser, type of information in letterhead, length of headline). For the few subjective variables (e.g., concreteness of donation purpose, color proportion), we conduct cross-checks among the coders to ensure inter-rater reliability for all of the data accumulated. For brevity, we do not explain each variable in detail here, but we provide an overview in Table 2. Table 2 also provides references to the previous direct mail design studies (Table 1) that have analyzed particular design variables in a similar way. As observed, many of the design variables included in our study have not been empirically examined in the prior research.

5.1.4. Common design characteristics across industries

In total, we record 36 design characteristics with 68 distinct design attributes across mail elements and design categories that are common across both industries. Some design characteristics contain multiple attributes, either representing different aspects of the respective characteristic (e.g., 3 different types of accentuations) or different degrees of implementation (e.g., the proportion of color in the supplement). For these, we distinguish between the mutually exclusive and the overlapping design attributes in Table 2. Most of these variables are binary, indicating whether a design characteristic or its attribute is observed (=1) in a campaign or not (=0). Only a few variables are metric, such as the length of headline or the number of pages in the supplement. We indicate these variables in Table 2. The frequencies or the means per industry and for the pooled data set are reported there.

5.1.5. Industry-specific variables

Based on the theoretical considerations, the extant literature, and the interviews with industry experts, we additionally record 21 industry-specific variables: 10 (with 31 design attributes) for the NPO and 11 (with 29 design attributes) for the FSP subsamples. These industry-specific design characteristics serve 2 purposes. First, the literature on direct marketing suggests that some of the effects of the direct mailing design characteristics are highly industry-specific (e.g., the type of testimonial for NPOs or awards and exemplary calculations for the FSP industry; Smith & Berger, 1996; Stone & Jacobs, 2008). Second, the design characteristics simultaneously act as controls: the expert interviews indicate that one-time donations require a different approach than continuous donation requests and hence the design of the mailing has to be adapted accordingly. Correspondingly, in the FSP industry, selling investment funds involves a different communication approach than selling consumer loans. Accounting for these differences helps to avoid biases when assessing the impact of the common design characteristics. Table 2 also contains these variables and the respective information.

5.1.6. Common controls

Based on the literature review and the expert interviews, we integrate 3 variables as common controls. First, the relative campaign volume (CV) is calculated by dividing the number of mailings per campaign by the total annual campaign volume in the NPO or FSP industries. Accordingly, the CV controls for the relative selectiveness of firms in choosing mail recipients in their campaigns (Bult & Wansbeek, 1995; Donkers et al., 2006). For example, target groups, and likewise CV, will usually be smaller if ambitious response goals and specific target groups guide the selection process. Second, we summate CVs for each NPO and FSP, resulting in the medium-specific relative annual sender volume (SV), i.e., reflecting the organization’s share of voice in the letter box of households. This variable accounts for the differences in share of voice, which are typically higher for larger organizations. Both controls, CV and SV, are also tested for nonlinear effects via squared terms (i.e., CV² and SV²). Third, we control for the average reception date within a month for all campaigns (1, 2, ..., 31), which was provided to us by GfK. Given the nature of how the data are collected (i.e., panel members send direct mailings to GfK at the end of the month), it is conceivable that keeping behavior could be higher for the direct mail received at the end of the month because consumers have less time to respond and therefore choose to keep it (the “end-of-month” effect). This variable reflects the number of days that have passed in a month. Hence, following this line of argument, the larger the number is, the higher the KR should be. Fourth, we add an effect-coded industry dummy to the pooled analysis to account for industry-specific effects (NPO = 1; FSP = −1).

5.2. Modeling and estimation approach

5.2.1. Model specification

Both dependent variables, the OR and the KR, are measured as fractions with a double truncation at 0 and 1. Accordingly, we employ a logit transformation to both variables to reduce their departures from non-normality (Ailawadi, Pauwels, & Steenkamp, 2008; Kraft, Albers, & Lal, 2004) and rename them LOR and LKR, respectively. Corresponding to our conceptual framework (cf. Fig. 1), we formulate regression equations for both dependent variables, LOR and LKR, for each sample i = 1, 2, 3 (NPO, FSP, and the pooled sample, respectively).
<table>
<thead>
<tr>
<th>Mail element</th>
<th>Type of design</th>
<th>Parameter</th>
<th>Characteristic</th>
<th>Operationalization</th>
<th>Selected sources</th>
<th>Obs. learning</th>
<th>Sample descriptives</th>
<th>Expected impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NPO (396 Mails)</td>
<td>FSP (281 Mails)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR</td>
<td>KR</td>
</tr>
<tr>
<td>Envelope (E)</td>
<td>VD</td>
<td>β</td>
<td>Weight</td>
<td>&gt;20 g</td>
<td>New (1)</td>
<td>Yes</td>
<td>94</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Format</td>
<td>Special and large sizes</td>
<td>6,7,10, (11,12,13)</td>
<td>Yes</td>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Presence of address window</td>
<td>No/yes</td>
<td>6,7 (12,13)</td>
<td>Yes</td>
<td>391</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Presence of teaser</td>
<td>No/yes</td>
<td>4,6,10 (11,12,13,14)</td>
<td>Yes</td>
<td>211</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Teaser content</td>
<td>Urgency appeal</td>
<td>New (1) (13,14)</td>
<td>Yes</td>
<td>57</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Questioning techniques</td>
<td>7 (11,13,14)</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>+ + 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sweepstake/freebie</td>
<td>No/yes</td>
<td>New (1) (11,13,14)</td>
<td>Yes</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Promotional design on back side</td>
<td>No/yes</td>
<td>6,7,10 (11,12,13)</td>
<td>Yes</td>
<td>286</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Promotional design on front side</td>
<td>No/yes</td>
<td>New (1) (11)</td>
<td>Yes</td>
<td>151</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sender's name on back side</td>
<td>No/yes</td>
<td>4,6,10 (11,12,13)</td>
<td>Yes</td>
<td>168</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sender's name on front side</td>
<td>No/yes</td>
<td>New (1) (11,12,13)</td>
<td>Yes</td>
<td>143</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Postage placement</td>
<td>Stamp</td>
<td>4,6,10 (11,12,13)</td>
<td>Yes</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Postage paid</td>
<td>No/yes</td>
<td>New (1) (11,12,13)</td>
<td>Yes</td>
<td>339</td>
<td>189</td>
</tr>
<tr>
<td>Letter (L)</td>
<td>VD</td>
<td>γ</td>
<td>Presence of headline</td>
<td>No/yes</td>
<td>New (1) (11,12)</td>
<td>Yes</td>
<td>215</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Length of headline (metric)</td>
<td>No. of words</td>
<td>New (1)</td>
<td>4.8</td>
<td>9.5</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Post scriptum</td>
<td>Presence no/yes</td>
<td>4,6,10 (11,12,13,14)</td>
<td>Yes</td>
<td>302</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Summary of offer</td>
<td>New (1)</td>
<td>4,5,6,10 (11,12,13)</td>
<td>Yes</td>
<td>203</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>New aspect/info</td>
<td>Pres. appeal</td>
<td>5 (11)</td>
<td>Yes</td>
<td>67</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Response appeal</td>
<td>No/yes</td>
<td>New (1) (12)</td>
<td>Yes</td>
<td>13</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Topography: letters with serifs</td>
<td>No/yes</td>
<td>New (1) (11,12)</td>
<td>Yes</td>
<td>104</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bold type</td>
<td>New (1)</td>
<td>5 (11)</td>
<td>Yes</td>
<td>134</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accentuations</td>
<td>Underlined</td>
<td>New (1) (11,12)</td>
<td>Yes</td>
<td>162</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Capital letters</td>
<td>New (1)</td>
<td>11 (11)</td>
<td>Yes</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Color of paper</td>
<td>White vs. colored</td>
<td>New (1)</td>
<td>6 (11)</td>
<td>16</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Color of font</td>
<td>Uniform black/blue vs. partly colored</td>
<td>New (1)</td>
<td>6 (11)</td>
<td>16</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Colored background print</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Colored illustration</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>281</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Content of letter head</td>
<td>Logo</td>
<td>New (1)</td>
<td>Yes</td>
<td>368</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Address</td>
<td>New (1)</td>
<td>Yes</td>
<td>280</td>
<td>201</td>
<td>481</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Phone number</td>
<td>New (1)</td>
<td>Yes</td>
<td>209</td>
<td>123</td>
<td>332</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fax number</td>
<td>New (1)</td>
<td>Yes</td>
<td>186</td>
<td>43</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Website</td>
<td>New (1)</td>
<td>Yes</td>
<td>171</td>
<td>60</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Photo of sender</td>
<td>New (1)</td>
<td>Yes</td>
<td>103</td>
<td>6</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E-mail</td>
<td>New (1)</td>
<td>Yes</td>
<td>136</td>
<td>14</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Toll-free phone number</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Presence of testimonial</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>8</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Presence of calendar date</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>350</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Current calendar date</td>
<td>No/yes</td>
<td>New (1)</td>
<td>Yes</td>
<td>350</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Presence of sender’s signature</td>
<td>No/yes</td>
<td>New (1) (11,12,13,14)</td>
<td>Yes</td>
<td>145</td>
<td>32</td>
</tr>
<tr>
<td>Length of letter</td>
<td>II</td>
<td>β</td>
<td>&gt; 1 page</td>
<td>New (1) (12,13,14)</td>
<td>Yes</td>
<td>2.6</td>
<td>145</td>
<td>32</td>
</tr>
<tr>
<td>NPO</td>
<td></td>
<td>γ</td>
<td>Tonalinty</td>
<td>Factual</td>
<td>New (1) (12)</td>
<td>Yes</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Emotional</td>
<td>New (1) (12)</td>
<td>Yes</td>
<td>324</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Concrete</td>
<td>New (1)</td>
<td>Yes</td>
<td>148</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vague</td>
<td>New (1)</td>
<td>Yes</td>
<td>248</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Doctor</td>
<td>New (1)</td>
<td>Yes</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Celebrity</td>
<td>New (1)</td>
<td>Yes</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Helper</td>
<td>New (1)</td>
<td>Yes</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Others</td>
<td>New (1)</td>
<td>Yes</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: OR = Other Research, KR = Known Research.
<table>
<thead>
<tr>
<th>Section</th>
<th>Code</th>
<th>Description</th>
<th>Yes/No</th>
<th>New Attributes</th>
<th>Value(s)</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSP</td>
<td>γ2,59</td>
<td>Time-limited offer</td>
<td>No/yes</td>
<td>New (1)</td>
<td>86</td>
<td>+/− n.a.; urgency</td>
</tr>
<tr>
<td></td>
<td>γ2,60</td>
<td>Offer details</td>
<td>No/yes</td>
<td>New (1)</td>
<td>43</td>
<td>+/− n.a.; information details</td>
</tr>
<tr>
<td></td>
<td>γ2,61</td>
<td>Restrictive terms and conditions (e.g., credit rating)</td>
<td>No/yes</td>
<td>New (1)</td>
<td>81</td>
<td>− n.a.; exclusion</td>
</tr>
<tr>
<td></td>
<td>γ2,62</td>
<td>Information on nearest branch</td>
<td>No/yes</td>
<td>New (1)</td>
<td>33</td>
<td>+/− n.a.; helpful information</td>
</tr>
<tr>
<td></td>
<td>γ2,63</td>
<td>Information on personal advisor/contact</td>
<td>No/yes</td>
<td>New (1)</td>
<td>37</td>
<td>+/− n.a.; helpful information</td>
</tr>
<tr>
<td>Supplement (S)</td>
<td>βS,114</td>
<td>Presence of Supplement</td>
<td>0% Colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(mutually exclusive attributes)</td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0% Colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(mutually exclusive attributes)</td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0% Colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(mutually exclusive attributes)</td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0% Colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(mutually exclusive attributes)</td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Up to 25% colored</td>
<td>New (1)</td>
<td>12</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26%–50% Colored</td>
<td>New (1)</td>
<td>17</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51%–75% Colored</td>
<td>New (1)</td>
<td>40</td>
<td>+/− curiosity; information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76%–100% Colored</td>
<td>New (1)</td>
<td>145</td>
<td>+/− curiosity; information</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Mail element</th>
<th>Type of variables</th>
<th>Characteristic</th>
<th>Operationalization(^b)</th>
<th>Selected sources(^c)</th>
<th>Sample descriptives(^b)</th>
<th>Expected impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offer (OFF)</td>
<td>NPO</td>
<td>γ(_{1,71})</td>
<td>Value appearance of give-away (mutually exclusive attributes)</td>
<td>Low</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,72})</td>
<td></td>
<td>Medium</td>
<td>4</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>FSP</td>
<td>γ(_{1,73})</td>
<td>Charitable category(^a)</td>
<td>Child aid</td>
<td>New (1)</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,74})</td>
<td></td>
<td>Diseases/disabilities</td>
<td>New (1)</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,75})</td>
<td></td>
<td>Environment/animals</td>
<td>New (1)</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,76})</td>
<td></td>
<td>Foreign aid</td>
<td>New (1)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,77})</td>
<td></td>
<td>Religion/church</td>
<td>New (1)</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,78})</td>
<td>Goal/intention of charitable mail(^a)</td>
<td>One-time donation</td>
<td>New (1)</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,79})</td>
<td></td>
<td>Continuous donations</td>
<td>New (1)</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,80})</td>
<td></td>
<td>Mere information</td>
<td>New (1)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,81})</td>
<td></td>
<td>Recruiting new members</td>
<td>New (1)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,82})</td>
<td></td>
<td>Thank-you letter</td>
<td>New (1)</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,83})</td>
<td>Adoption/sponsorship</td>
<td>New (1)</td>
<td>8</td>
<td>− n.a.; higher commitment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,77})</td>
<td>Product category(^a)</td>
<td>Loans</td>
<td>New (1)</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,79})</td>
<td></td>
<td>Savings/investments</td>
<td>New (1)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,80})</td>
<td></td>
<td>Stocks/funds</td>
<td>New (1)</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,81})</td>
<td></td>
<td>Credit card</td>
<td>New (1)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,82})</td>
<td></td>
<td>Retirement provisions</td>
<td>New (1)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,83})</td>
<td></td>
<td>Home purchase savings</td>
<td>New (1)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,84})</td>
<td></td>
<td>Information, no offer</td>
<td>New (1)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>γ(_{1,85})</td>
<td></td>
<td>Investment advice</td>
<td>New (1)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−</td>
<td></td>
<td>Others</td>
<td>New (1)</td>
<td>27</td>
</tr>
<tr>
<td>Opening rate</td>
<td>Common Controls</td>
<td>β(_{1,58})</td>
<td>Campaign Volume (metric) (main effect and squared)</td>
<td>New (1)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β(_{1,59})</td>
<td>Firm Volume (metric) (main effect and squared)</td>
<td>New (1)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β(_{1,62})</td>
<td>“End of Month”-Effect (metric)</td>
<td>New (1)</td>
<td>17.6</td>
<td>14.6</td>
</tr>
</tbody>
</table>

\(\text{VD}=\text{visual design, SI=sender identity, P=personalization, II=information intensity, NPO/FSP=industry-specific design variable; n.a.}=\text{not applicable.}\)

\(^a\) Record by GfK.

\(^b\) Item frequency for dummy variables (yes) and means for metric variables.

\(^c\) Sources: 1 New, based on theory, expert interviews and empirical assessments; Empirical Studies (see Table 1): 2 Beard et al. (1990); 3 Bekkers and Crutzen (2007); 4 Bell et al. (2006); 5 Bult et al. (1997); 6 De Wulf et al. (2000); 7 James and Li (1993); 8 Ledolter and Sweesey (2006); 9 Van der Scheer et al. (1996); 10 Vriens et al. (1998); Textbooks: 11 Geller (2002); 12 Jones (1997); 13 Nash (1997); 13 Nash (2000); 14 Roberts and Berger (1999); 15 Stone and Jacobs (2008)—Textbook sources are no (refereed) empirical studies.
In each of these 3 samples, we have a different number of \( j(i) = 1, 2, \ldots, f(i) \) campaigns (\( j(1) = 396, j(2) = 281, j(3) = 677 \), respectively). To analyze the effects of \( K(i) \) sample-specific independent variables on the respective dependents (see Table 2 for details), we employ an OLS regression on the respective equations specified in Eqs. (2) and (3):

\[
\log_{ij} = \beta_{0j} + \sum_{k=1}^{K(i)} \beta_{jk}x_{ikj} + u_{ij}
\]

(2)

In Eq. (2), \( \log_{ij} \) refers to the dependent variable, the logit of the OR, in sample \( i \) for campaign \( j \). Across campaigns, the dependent variable is explained by a sample-specific set of parameters (\( \beta_{jk} \)) and a corresponding set of explanatory variables (\( x_{ikj} \)), resulting in a normally distributed error (\( u_{ij} \)). Across all samples, the parameters include an intercept (\( \beta_{0j} \)) and 11 design characteristics for the envelope with 15 design attributes, of which one is defined by the other 2 exclusive alternatives (\( \beta_{11j}, \beta_{114} \); see Table 2). Additionally, we incorporate several design characteristics that indicate observational learning, where the recipients might sense special content in the direct mail piece. Apart from weight and format (\( \beta_{11}, \beta_{22} \)), which are already subsumed under the envelope characteristics, we associate letter length and the presence of a supplement or a response action with observational learning (\( \beta_{115j}, \beta_{117j} \)). The presence of a give-away is only relevant for the NPO industry (\( \beta_{112j} \)). Moreover, the controls CV and SV with their respective squared terms for testing a potentially nonlinear influence as well as the “end-of-month” effect enter the equation to control for the firms’ selectivity in choosing mail recipients and firm brand effects (\( \beta_{116j}, \beta_{1122} \)). For the pooled analysis, we add an effect-coded industry dummy (\( \beta_{122} \)) to account for the industry-specific effects.

\[
\log_{ij} = \gamma_{0i} + \sum_{k=1}^{K(i)} \gamma_{ik}x_{ik} + v_{ij}
\]

(3)

In Eq. (3), \( \log_{ij} \) refers to the dependent variable, the logit of KR, in sample \( i \) for campaign \( j \). Across campaigns, the dependent variable is explained by a sample-specific set of parameters (\( \gamma_{ik} \)) and a corresponding set of explanatory variables (\( z_{ik} \)), resulting in a normally distributed error (\( v_{ij} \)). Across all samples, the parameters include an intercept (\( \gamma_{0i} \)) and 15 design characteristics for the letter with 27 design level attributes (\( \gamma_{21i}, \gamma_{27i} \)), 4 characteristics for the supplement with 8 design level attributes (\( \gamma_{28i}, \gamma_{36i} \)) and 4 characteristics for the response device with 16 design level attributes (\( \gamma_{52}, \gamma_{52} \)). Analogous to Eq. (2), we account for the potential (nonlinear) effects of CV and SV as well as for the “end-of-month” effect (\( \gamma_{53}, \gamma_{57} \)). To test the link between the 2 stages, we include OR in both industries (\( \gamma_{0i} \)). For the pooled analysis, we again add an industry dummy (\( \gamma_{159} \)) to account for the industry-specific effects.

In the NPO (FSP) model, we extend the \( Z_{1} (Z_{2}) \) vector by the 10 (11) industry-specific variables from Table 2 (\( \gamma_{159}, \gamma_{183}; \gamma_{259}, \gamma_{285} \), respectively).

To limit the industry sample size effects in the pooled analysis, we weight all cases from the NPO (FSP) industry with a factor of .8548 (1.2046).

5 As some design level attributes are mutually exclusive, they require 1 parameter less (see Table 2 for details).

5.2.3. Robustness checks

We specifically test for linear model assumptions. To assess the degree of multicollinearity, we calculate the bivariate correlations and the variance inflation factors (VIFs) (see the correlation matrices and the VIFs in Appendix). In the OR models, all of the VIF scores are below 6 and thus do not exceed the critical values (Belsley, Kuh, & Welsh, 2004). For instance, the highest VIF scores for the full OR model in the NPO industry are 3.609 and 3.102 for campaign volume and its squared equivalent, respectively; in the FSP industry, they are 5.893 and 5.743 for firm volume and its squared equivalent, respectively. Both VIF values are somewhat lower in the final model, as can be expected after variable elimination. For the full KR models, we initially find substantially higher VIFs for the 12 variables in each industry (e.g., the VIFs for presence and type of response device were 56.478 and 46.320, respectively, for NPOs; the VIFs for presence and color proportion of the brochure 51–75% were 213.616 and 190.888, respectively, for FSPs). However, in the final KR models, these variables are eliminated and the VIFs of all of the remaining variables fall well below the critical values.6 In the final models, the correlations between the variables are also relatively low, with the vast majority of correlation coefficients below .3. Thus, multicollinearity does not affect our final results. Both the Breusch–Pagan–Godfrey test and the White test indicate a substantial level of heteroscedasticity. Hence, we correct for heteroscedasticity by applying White’s (1980) correction to derive robust standard errors. Furthermore, we control for correlated error terms between the OR and the KR model by applying Breusch and Pagan’s Lagrange Multiplier test (Breusch & Pagan, 1980). The correlation between the residuals of our final OR and KR models are below .10 for both FSPs and NPOs, resulting in insignificant Lagrange multipliers (p-value NPO = .63; p-value FSP = .36). Accordingly, the OR and KR equations can be estimated independently. Nevertheless, we test our final estimation results for the OR and the KR by also running a Seemingly Unrelated Regression (Zellner, 1962) and find no distinct results with regard to both the relative effects of the variables and their significance levels. With the OR being the dependent variable in Eq. (2) and a potential predictor in Eq. (3), we also conducted Hausman’s (1978) residual test to account for any potential endogeneity problems caused by the simultaneity of the OR in the first estimation step. The test results do not indicate any need to modify our models and our estimation approach in the first step, and the OR is subsequently dropped in the second step for both industries.

6 We selectively included variables with high VIF scores and t-values lower than 1 in the final models but were unable to detect additional significant effects for these variables. Hence, we did not extend our final models further.

6. Empirical results

6.1. Direct mail funnel and the impact of design characteristics

First, we report the findings on the connection between the 2 stages of the direct mail funnel as well as the impact of the design characteristics at those stages, represented by the OR and KR, respectively. Concerning the relationship between the OR and the KR of campaigns, the positive spill-over effect presumed in our conceptual model (see Fig. 1) is not confirmed. In neither of the 2 industries do we find a significant effect for the OR on the KR. With t-values of –.29 (NPO) and .99 (FSP), the OR was dropped from the KR model in the initial stage of the variable selection process because it had t-values smaller than 1. This finding implies that a higher/lower OR does not imply a higher or lower KR. Both rates are statistically independent of each other for both industries.

In Table 3, we report the estimation results for the first funnel stage for the 2 industries (see Eq. (2)) regarding the common and

(e.g., Dekimpe & Hanssens, 1995; Kraft et al., 2004; Pesaran, Pierse, & Lee, 1993).
industry-specific design effects on the OR. We also report a pooled analysis across the 2 industries, where the samples are weighted by size and only the effects that are significant at the 5-percent level (2-sided, \( \tau = 1.97 \)) in either of the 2 industries are kept. The overall goodness-of-fit criteria indicate a reasonable explanatory power for our parsimonious models for the 2 industries with adj. \( R^2 \) s of .173 for NPOs and .260 for FSPs. Table 4 shows the estimations per industry and the weighted pooled analysis for the KR. The adjusted \( R^2 \) for the NPO is .340, while it is lower for FSPs with a value of .164. For both the OR and the KR equations, the adjusted \( R^2 \) s are lower or similar for the weighted pooled analysis, with values of .116 (OR) and .173 (KR).

### 6.2. Impact of design characteristics by category

We base our assessments of the effects of each design characteristic on the \( t \)-statistics (see Tables 3 and 4). In our discussion, we focus on the effects with a \( p \)-value < .05.

#### 6.2.1. Visual design category

Regarding the OR, a colored envelope design for the envelope has negative main effects on the OR for both industries. This finding is confirmed in the pooled analysis across both samples. The industry-specific estimates indicate that larger format sizes for the envelope, teasers employing the questioning technique and a promotional design on the backside of the envelope show positive main effects in the NPO industry. However, only the positive effect of a teaser with a questioning technique is confirmed in the pooled analysis. Interestingly, special and larger envelope designs positively influence the OR in the NPO industry, while exerting a negative, weakly significant effect in the FSP industry. In the FSP industry, a larger weight positively influences the OR while the presence of a teaser reveals a negative main effect. These effects are confirmed in the pooled analysis. Overall, most of the significant visual design variables in the industry-specific analyses are also significant in the pooled analysis.

With respect to the KR, for the NPO industry, only the pre-stamped envelope as a response device yields a positive influence. In the FSP industry we do, however, find 3 significant main effects. The length of the headline is positively related to the KR. The presence of a post-scripturn negatively influences the KR. However, when this post-scripturn provides some new information, it positively influences the KR. In the pooled model, the latter variable is the only significant visual design characteristic with a positive influence on the KR.

#### 6.2.2. Sender identity

Concerning the OR, we do not find any significant effect for this type of design characteristic in the NPO industry. In the FSP industry, placing the sender’s name on the front side has a negative main effect. In the pooled analysis, this effect remains negative and significant.

For the KR, our estimation results in both industries show positive effects for placing the company logo and the fax number in the letterhead. The effect of the logo is confirmed in the pooled model; however, the fax number is only marginally significant. Remarkably, the presence of a phone number for the sender in the letterhead negatively influences the KR in the NPO industry. One reason could be that the recipients expect a toll-free number.

#### 6.2.3. Personalization category

At the OR stage, envelope personalization is captured by the type of postage placement. Using the least personalized option, i.e., the imprint “postage paid,” leads to significantly lower overall ORs in the FSP industry. This negative main effect is also found in the pooled analysis.

With respect to the KR, the personalization of the supplement exerts a positive main effect in both industries. Although a personalized supplement is significantly positive in the pooled model and for the NPO industry, it is only marginally significant in the FSP industry.

#### 6.2.4. Information intensity category

No significant variables capturing the information intensity of the mailing are found in the OR equations across industries. This result is hardly surprising because the envelope usually does not contain substantially varying degrees of information.

The aspects that reflect the information intensity of the direct mail package are captured for the KR. Here, the length of the cover letter exerts a positive influence in the NPO industry, while the length of the brochure has a positive main effect in the FSP industry. In the pooled model, only the positive effect of the letter length is confirmed.
### Table 4

Empirical results on keeping rates.

<table>
<thead>
<tr>
<th>Variables</th>
<th>NPO Parameter Estimate</th>
<th>NPO t-Value</th>
<th>FSP Parameter Estimate</th>
<th>FSP t-Value</th>
<th>Weighted pooled Parameter Estimate</th>
<th>Weighted pooled t-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>γ2.0</td>
<td>-5.854 (.729)</td>
<td>8.00***</td>
<td>γ2.0</td>
<td>-6.777 (.583)</td>
<td>-11.62***</td>
</tr>
<tr>
<td>Visual design:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L: Length of headline</td>
<td>γ2.2</td>
<td>1.083 (.840)</td>
<td>1.29</td>
<td>γ2.2</td>
<td>1.045 (.430)</td>
<td>2.37***</td>
</tr>
<tr>
<td>L: Presence of post scriptum</td>
<td>γ2.3</td>
<td>1.277 (.554)</td>
<td>2.22***</td>
<td>γ2.3</td>
<td>-1.07 (.02)</td>
<td>-1.07</td>
</tr>
<tr>
<td>L: Post scriptum: new aspect/info</td>
<td>γ2.4</td>
<td>.363 (.442)</td>
<td>.81</td>
<td>γ2.4</td>
<td>- .72 (.08)</td>
<td>-1.36</td>
</tr>
<tr>
<td>L: Typography: letters with serifs</td>
<td>γ2.5</td>
<td>-1.01 (.395)</td>
<td>-.56***</td>
<td>γ2.5</td>
<td>- 1.31 (.58)</td>
<td>-2.11***</td>
</tr>
<tr>
<td>L: Accentuations: capital letters</td>
<td>γ2.6</td>
<td>- .507 (.267)</td>
<td>-1.90</td>
<td>γ2.6</td>
<td>-1.163 (.576)</td>
<td>-2.8**</td>
</tr>
<tr>
<td>L: Color of paper</td>
<td>γ2.7</td>
<td>.122 (.315)</td>
<td>3.90***</td>
<td>γ2.7</td>
<td>1.399 (.466)</td>
<td>3.00***</td>
</tr>
<tr>
<td>RD: Pre-stamped envelope</td>
<td>γ2.8</td>
<td>-1.010 (.395)</td>
<td>-2.56***</td>
<td>γ2.8</td>
<td>- .32 (.412)</td>
<td>-0.78</td>
</tr>
<tr>
<td>L: Content of letter head: logo</td>
<td>γ2.9</td>
<td>1.506 (.396)</td>
<td>3.80***</td>
<td>γ2.9</td>
<td>1.094 (.443)</td>
<td>2.47***</td>
</tr>
<tr>
<td>L: Content of letter head: phone number</td>
<td>γ3.0</td>
<td>- .507 (.267)</td>
<td>-1.90</td>
<td>γ3.0</td>
<td>-1.163 (.576)</td>
<td>-2.8**</td>
</tr>
<tr>
<td>S: Length of brochure</td>
<td>γ3.1</td>
<td>.635 (.242)</td>
<td>2.62***</td>
<td>γ3.1</td>
<td>.521 (.489)</td>
<td>1.05</td>
</tr>
<tr>
<td>L: Type of testimonials: doctor</td>
<td>γ3.2</td>
<td>- .398 (.521)</td>
<td>.77</td>
<td>γ3.2</td>
<td>.731 (.354)</td>
<td>2.07***</td>
</tr>
<tr>
<td>L: Type of testimonials: helper</td>
<td>γ3.3</td>
<td>2.430 (1.029)</td>
<td>2.36***</td>
<td>γ3.3</td>
<td>.843 (.440)</td>
<td>1.90</td>
</tr>
<tr>
<td>L: Kind of give-away: calendar</td>
<td>γ3.4</td>
<td>.511 (.516)</td>
<td>.99</td>
<td>γ3.4</td>
<td>1.181 (.309)</td>
<td>5.87***</td>
</tr>
<tr>
<td>L: Kind of give-away: others</td>
<td>γ3.5</td>
<td>.699 (.360)</td>
<td>1.69</td>
<td>γ3.5</td>
<td>.890 (.461)</td>
<td>1.93</td>
</tr>
<tr>
<td>RC: Kind of give-away: high</td>
<td>γ3.6</td>
<td>-1.010 (.395)</td>
<td>-2.56***</td>
<td>γ3.6</td>
<td>- 1.31 (.58)</td>
<td>-2.8**</td>
</tr>
<tr>
<td>OFF: Charitable category:</td>
<td>γ3.7</td>
<td>-.553 (.324)</td>
<td>-1.71</td>
<td>γ3.7</td>
<td>.520 (.471)</td>
<td>1.11</td>
</tr>
<tr>
<td>Information intensity:</td>
<td>γ3.8</td>
<td>-1.942 (.327)</td>
<td>-5.94***</td>
<td>γ3.8</td>
<td>.578 (.228)</td>
<td>2.54***</td>
</tr>
<tr>
<td>Industry-specific variables (NPO):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L: Type of testimonials: doctor</td>
<td>γ3.9</td>
<td>.475 (.561)</td>
<td>.85</td>
<td>γ3.9</td>
<td>.475 (.561)</td>
<td>.85</td>
</tr>
<tr>
<td>L: Type of testimonials: helper</td>
<td>γ4.0</td>
<td>2.430 (1.029)</td>
<td>2.36***</td>
<td>γ4.0</td>
<td>2.430 (1.029)</td>
<td>2.36***</td>
</tr>
<tr>
<td>L: Kind of give-away: calendar</td>
<td>γ4.1</td>
<td>.511 (.516)</td>
<td>.99</td>
<td>γ4.1</td>
<td>.511 (.516)</td>
<td>.99</td>
</tr>
<tr>
<td>L: Kind of give-away: others</td>
<td>γ4.2</td>
<td>.699 (.360)</td>
<td>1.69</td>
<td>γ4.2</td>
<td>.699 (.360)</td>
<td>1.69</td>
</tr>
<tr>
<td>L: Kind of give-away: high</td>
<td>γ4.3</td>
<td>-1.010 (.395)</td>
<td>-2.56***</td>
<td>γ4.3</td>
<td>- 1.010 (.395)</td>
<td>-2.56***</td>
</tr>
<tr>
<td>OFF: Charitable category:</td>
<td>γ4.4</td>
<td>-.553 (.324)</td>
<td>-1.71</td>
<td>γ4.4</td>
<td>-.553 (.324)</td>
<td>-1.71</td>
</tr>
<tr>
<td>OFF: Charitable category:</td>
<td>γ4.5</td>
<td>-1.942 (.327)</td>
<td>-5.94***</td>
<td>γ4.5</td>
<td>-1.942 (.327)</td>
<td>-5.94***</td>
</tr>
<tr>
<td>OFF: Goal/intention of charitable mail:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC: Kind of give-away: high</td>
<td>γ4.6</td>
<td>-.553 (.324)</td>
<td>-1.71</td>
<td>γ4.6</td>
<td>-.553 (.324)</td>
<td>-1.71</td>
</tr>
<tr>
<td>OFF: Goal/intention of charitable mail:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry-specific variables (FSP):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L: Offer details</td>
<td>γ5.0</td>
<td>.979 (.418)</td>
<td>2.34***</td>
<td>γ5.0</td>
<td>.979 (.418)</td>
<td>2.34***</td>
</tr>
<tr>
<td>L: Restrictive terms and conditions</td>
<td>γ5.1</td>
<td>- .731 (.354)</td>
<td>-2.07**</td>
<td>γ5.1</td>
<td>- .731 (.354)</td>
<td>-2.07**</td>
</tr>
<tr>
<td>RD: Response options: fill out application form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD: Response options: request additional information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD: Information required from recipient: mailing address</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campaign volume (linear)</td>
<td>γ5.3</td>
<td>1.765 (.241)</td>
<td>7.32***</td>
<td>γ5.3</td>
<td>1.765 (.341)</td>
<td>7.32***</td>
</tr>
<tr>
<td>Campaign volume (squared)</td>
<td>γ5.4</td>
<td>- .464 (.065)</td>
<td>-4.70***</td>
<td>γ5.4</td>
<td>- .464 (.065)</td>
<td>-4.70***</td>
</tr>
<tr>
<td>Firm volume (linear)</td>
<td>γ5.5</td>
<td>.424 (.196)</td>
<td>2.17**</td>
<td>γ5.5</td>
<td>.359 (.167)</td>
<td>2.17**</td>
</tr>
<tr>
<td>Firm volume (squared)</td>
<td>γ5.6</td>
<td>- .893 (.184)</td>
<td>-4.84***</td>
<td>γ5.6</td>
<td>- .893 (.184)</td>
<td>-4.84***</td>
</tr>
<tr>
<td>&quot;End of month&quot;-effect</td>
<td>γ5.7</td>
<td>.031 (.014)</td>
<td>2.12**</td>
<td>γ5.7</td>
<td>.031 (.014)</td>
<td>2.12**</td>
</tr>
<tr>
<td>Industry dummy</td>
<td>γ5.8</td>
<td>.340 (.164)</td>
<td>2.07</td>
<td>γ5.8</td>
<td>.355 (.132)</td>
<td>2.69***</td>
</tr>
<tr>
<td>Adj. R²</td>
<td></td>
<td>.792***</td>
<td>3.619***</td>
<td></td>
<td></td>
<td>8.607***</td>
</tr>
</tbody>
</table>

a. OLS-Model estimated with White’s (1980) correction for heteroscedasticity (robust standard errors).

**  sign. at p<.01.

***  sign. at p<.005.

### 6.2.5. Observational learning

Across all design categories, we identify those design characteristics that allow the recipient to learn about the expected content of the direct mail piece. Because these types of characteristics potentially generate curiosity, they might positively influence the OR. In

Table 2, all of these design characteristics are identified in the respective column. Across both industries, we find that the larger formats in the NPO industry and a higher weight in the FSP industry raise the OR in these industries. In the pooled sample, only a higher weight is confirmed as driving the OR.
6.2.6. Industry-specific design characteristics

We do not capture industry-specific design characteristics for the OR stage. For the KR in the NPO industry, depicting a helper as a testimonial enhances the KR. In contrast, the campaigns with a high-value appearance for the giveaway, campaigns from religious institutions, or campaigns with the primary objective of recruiting new members lead to lower KRs. For the FSPs, we find that presenting offer details in the letter or offering the response option to request additional information exerts a positive effect on the KR. Displaying restrictive terms and conditions or asking the recipient to fill out an application form leads to significantly lower KRs. This latter finding parallels with the new membership request in the NPO industry, suggesting that any initial direct mail solicitations in both industries should not be intrusive.

6.2.7. Common controls

At the OR stage of the direct mail funnel, the results show that the ORs drop with increasing campaign mailing volumes in the NPO industry only. This drop, however, becomes smaller as the campaign volume increases because we find a positive significant quadratic term for the campaign volume (U-shaped relationship). In the FSP industry, no significant effects for the campaign volume are found. The firm volume appears to be positively related to the OR in the NPO industry, while its main effect is negative in the FSP industry. The positive quadratic coefficient for the firm volume in the FSP industry suggests that this negative effect becomes smaller as the firm volume increases (U-shape).

Overall, our results suggest that there is mixed evidence between the 2 industries on the role of the campaign and the firm volume on the OR. As none of the effects can be found in the pooled sample, it appears that these effects are rather industry-specific.

At the KR stage of the direct mail funnel, we detect a curvilinear, inverted U-shape relationship between the campaign volume and the KR and between the firm volume and the KR in the NPO industry, given that the parameter estimate for the linear (quadratic) term is significant and positive (negative). These effects are absent in the FSP industry. However, both effects are also found in the pooled model, but the quadratic effect for the campaign volume is only weakly significant here. We also observe a positive end-of-month effect in the NPO industry, suggesting that there are higher KRs when the mailings are sent at the end of the month. This positive effect can be detected only marginally in the pooled model. The industry dummy is only significant in the pooled analysis for the KR, indicating that there are higher overall KRs in the NPO industry.

7. Discussion

7.1. Insights into the direct mail funnel

The unique data set utilized in this study reveals that the previously unobservable “black box” of the intermediate stages of the direct mail funnel appears to contain valuable diagnostic information for marketers. This finding compares to the systematic use of intermediate communication measures for other media (such as awareness, recall, or recognition).

7.1.1. Industry differences in opening and keeping rates

Overall, the ORs are statistically similar in both industries, but non-profit organizations manage to attain higher overall KRs (indicated by a significant positive industry dummy, see Table 4). By habit or out of curiosity, consumers are generally inclined to open direct mail envelopes. In addition, when receiving a mailing from an FSP, many recipients are likely to open the envelope to make sure that no potentially important information is being missed, e.g., to confirm that they are not mistakenly discarding personal balance information. Additionally, the negative effect of a colored envelope design is somewhat more negative in the NPO industry and, at the same time, more widely used (75% vs. 35%, as observed in Table 2). Both effects combined might explain the marginally higher ORs in the FSP campaigns. Conversely, one conceivable explanation for the higher KRs in the NPO campaigns is that charitable solicitations, by their very nature, usually have a broader appeal (e.g., helping people) than financial offerings. The latter is typically more specific in nature (e.g., signing up for another credit card). Moreover, compared to other product categories, the recipients of non-profit mailings often exhibit a higher level of involvement with the charitable solicitation because of the emotional importance attached to helping behavior (e.g., Francis & Holland, 1999).

7.1.2. Disconnect between the opening and the keeping rate

Another noteworthy finding is the lack of a significant relationship between the ORs and the KRs, i.e., from opening the direct mail to keeping it. This finding suggests that there are no significant spill-over effects in the direct mail funnel at the aggregate campaign level, at least within our sample representing 2 major direct mail industries. In other words, opening a direct mail piece is obviously a necessary condition for responding to the offer, but it is not per se a driver of direct mail keeping and, accordingly, the subsequent response. As a consequence, both stages in the response process should be optimized independently.

It appears that curiosity in the initial contact stage induces consumers to look inside the envelope, but afterwards, benefit motives prevail. In some cases, the outside envelope might even have been misleading, resulting in dissonant feelings on the part of the recipient. Discarding the direct mail piece might then be used as a dissonance-reducing strategy (Festinger, 1957).

7.2. Impact of design characteristics on the opening and keeping rates

7.2.1. Overall impact of design characteristics

The direct mail design characteristics determine the campaign effectiveness to a substantial degree: for the NPOs (FSPs), they explain 13.7% (24.4%) of the total variance of the ORs and 21.5% (16.4%) of the KRs (i.e., when controls are dropped from the equation). Hence, our study furnishes the first empirical support for the claim offered in direct marketing textbooks suggesting that 10 to 25% of direct mail campaign success can be attributed to creative execution (e.g., Roberts & Berger, 1999, p. 7; Stone & Jacobs, 2008, p. 6).

7.2.2. Impact of specific mailing design categories

Our results indicate that the different categories of direct mail characteristics are of differential importance in explaining the ORs and the KRs. Whereas the visual design elements on the outer envelope appear to be the predominant drivers of the ORs, the other categories of design characteristics become comparatively more important for explaining the KRs. Notably, for the OR, the effects of several visual design characteristics differ between the 2 industries. This finding underlines the claims in the literature that the effectiveness of direct mail design is to a considerable extent industry-specific (e.g., Stone & Jacobs, 2008). Additionally, whereas 3 visual design variables exert a significant influence on the KRs in the FSP industry, only the pre-stamped response device drives the KR in the NPO industry.

Considering the effects of all visual design characteristics on the OR in concert, it appears that the FSPs are well advised to use plain envelopes resembling official business mail and to avoid design elements that signal the promotional quality of the mail (e.g., extraordinary formats, teasers, and colorful design). For the NPO mailings, a more nuanced picture emerges: while colorful envelope design should be avoided in most cases, some design features such as special envelope formats, teasers with questioning techniques or promotional designs on the back side can help to enhance the opening behavior. With respect to the sender identity-related variables, our findings are somewhat ambivalent. The ORs can be increased by withholding the sender’s name on the envelope. This tactic of creating curiosity is particularly effective in the FSP campaigns, but generates no
significant effect in the NPO campaigns. For the letter, however, our results unanimously show that placing the company’s logo in the letterhead can significantly enhance the KRs in both industries. This finding underscores the value of brand elements in marketing communications (e.g., Keller, 2007; Pieters & Wedel, 2004).

Our results regarding personalization are somewhat mixed. On the outer envelope, a personalization impression can be conveyed through the type of postage payment employed. In the FSP industry, the direct mail envelope should not appear to the recipient to be bulk mail. This impression could be caused by using a “posting paid” imprint instead of ink or real stamps, resulting in reduced ORs for the campaigns. Additionally, among the various personalization options for the other components of the mailing, only supplement personalization significantly and consistently enhances the KRs in both industries studied. Because all mailings in our data set are personally addressed by default, the vast majority of them also exhibit the standard personalization features such as a personal salutatory address, the current calendar date, the sender’s signature and a response device with the addressee’s name, etc. As a consequence, these personalization features are rather static and, thus, cannot exert differential effects across campaigns. By contrast, the personalization of the supplement is only used by relatively few companies in our sample and can thus serve as a differentiating factor. Hence, the personalization of specific elements can be an effective tactic compared to using standardized mailings. In sum, our results show that personalization is primarily a driver of the KR, thereby confirming the literature on survey response rates, which states that personalization is important (e.g., Dillman, 2007).

Interestingly, we find that 2 information intensity variables are positively related to the KRs. These variables are, however, different for the 2 industries. While the letter length positively influences the KR in the NPO industry, the length of the brochure positively influences the KR in the FSP industry. The finding of these 2 positive effects is noteworthy because it contrasts with the widely accepted notion of information overload in advertising and consumer behavior (e.g., Hoyer & Maclnnis, 2007). While this view has intuitive appeal, it should be noted that direct mail advertising usually receives deliberate and active attention by consumers once the envelope has been opened. In addition, not all information must be processed at the moment of opening the direct mail piece. If there is a brochure or a long letter in the envelope, it could well be that people keep the mail piece to read all of the information when they have the time. As a result, information overload due to the limited capacity of short-term memory is of less concern than it is in the case in real-time media such as TV or online advertising.

With respect to our controls, the effects of the campaign- and sender-level mailing volumes differ across industries and funnel stages. Again, this difference underlines the diagnostic value of analyzing the intermediate stages of the direct mail funnel. For the NPO campaigns, we find a positive relationship between the firm volume and the OR. In other words, the direct mail from large and well-known non-profit organizations has a higher likelihood of being opened. We also find a U-shaped relationship between the campaign volume and the ORs; up to some point, the ORs tend to drop with the increasing campaign volumes, suggesting that there are wastage effects due to less selective targeting. For very large campaigns, however, the ORs tend to increase again. This result could be because large campaigns are typically accompanied by cross-media support (e.g., web, TV, and radio advertising), and could thus benefit from heightened awareness.

For the KRs of NPO campaigns, we find a pronounced inverted U-shape relationship with both campaign volume and firm volume. On the one hand, the KRs start to drop beyond some optimal volume—most likely also due to a less targeted address selection. As large-volume campaigns inevitably address the less responsive consumer segments, the wastage effects drive down the KRs. On the other hand, it appears that charitable campaigns must reach some reasonable size and come from bigger organizations to be perceived by the recipient as trustworthy and relevant. Up to some point, at which the wastage effects start to dominate, it appears that trust and the positive image created through the brand name of a large and well-known non-profit organization enhance the effectiveness of charitable solicitations (Bendapudi, Singh, & Bendapudi, 1996). With regard to the timing of campaigns, we detect a significant “end of month” effect for the KR in the NPO sample, in that the direct mail pieces are more likely to be kept by the panel participants if the mail is received closer to the end of the month. Accordingly, if researchers and practitioners analyze this type of panel data, they need to control for this effect.

Compared to the NPO campaigns, the control variable effects for the FSPs appear to be less pronounced. We do not find any significant volume or timing effects on the KRs. In contrast to NPOs, we do not detect the positive linear effect of firm volume on the OR but find a U-shaped effect. It appears that the small and large FSPs (in terms of mailing volume) are better able than the medium-sized FSPs to entice consumers to examine their offers. For the large FSPs, the positive familiarity effect of well-known institutions is likely to play out. Small financial service providers, on the other hand, could be targeting very well-defined segments with direct mail pieces that signal exclusive offerings (e.g., regional players or private banks).

8. Implications

We show how researchers and firms can systematically investigate the effect of design characteristics on direct mail performance. Researchers can use the previously unavailable panel data on the intermediate stages of the direct mail funnel to better investigate heterogeneous effects across the funnel. The panel data also allow them to compare differences at the intermediate communication stages. Managers can leverage the commercially available direct mail panel data to augment their managerial tool set by covering a previous blind spot, namely the active management of design for direct mailings. Our approach in combination with the panel data offers them an opportunity to further improve the design of mailing characteristics as an important means to increase the OR and the KR as drivers of response rates.

Moreover, our study findings offer some specific guidelines for the marketing managers who are responsible for running direct mail campaigns. Researchers and managers are likely to gain new insights, as several of our design recommendations are in marked contrast with the current methods used to design the majority of direct mailings by companies. Even some non-significant or negative findings might be worth noting: for example, as shown in Table 2, 43% of all campaigns from NPOs (172 out of 396) contain some type of give-away to potential donors. However, the presence of giveaways does not appear to enhance the response process. To the contrary, our results suggest that the giveaways with a high-value appearance even lead to lower KRs, thus casting doubt on the benefits from the costly inclusion of giveaways. Likewise, over two thirds of direct mail letters from FSPs (198 out of 281) contain some type of postscript, as this is commonly assumed to be an effective technique. Our results challenge this common practice and paint a more nuanced picture: on average, the use of postscripts is associated with lower KRs in the FSP industry unless some new aspect or information regarding the offer is presented (see Table 4). Other features such as attaching payment devices to cover letters or the depiction of awards (e.g., “rated as best investment fund”) have been proposed to stimulate the response process but fail to exert significant effects in our study. As these and some other examples given below reveal, several of our findings are rather unexpected.

Our guidelines for managers can be summarized in some “direct mail design recommendations” to increase the intermediate performance of direct mail campaigns. Based on our study, the generalizable suggestions across campaigns and industries include the following.

- Use color with caution. This recommendation is based on our finding that a colorful envelope design reduces the ORs across both

S. Feld et al. / Intern. J. of Research in Marketing 30 (2013) 143–159

157
industries. As shown in Table 2, colored envelopes are actually employed in the majority of campaigns (i.e., in 395 out of 677 campaigns), but our results indicate that this prevailing practice might actually be counterproductive.

- Use your sender identity with care. Our results indicate that eliminating your name from the envelope can facilitate the ORs, particularly in FSP campaigns. For the letter, by contrast, the findings unanimously show that the KRs are higher if the letterhead contains a company logo. Hence, the direct marketers should capitalize on the positive brand communication effects in the letter to signal familiarity and gain trust. To establish this type of qualified contact, however, it might actually be beneficial to create curiosity by not identifying the sender of the promotional message up front on the envelope. Analogous to a personal selling situation, this recommendation is similar to getting a “foot in the door” as a necessary condition for presenting the offer.

- Provide sufficient information. In both industries, we find some evidence that providing more information increases the KR. NPOs should use long rather than short letters to convey enough information. Providing information is important for non-profit organizations because the prospective donors must first believe the charity’s message depicting need (Bendapudi et al., 1996). Longer texts could be helpful to present a variety of details on the non-profit organization itself as well as regarding the cause of the need and the objective of the corresponding donation. Similarly, offering comprehensive and detailed information is of paramount importance for financial service providers to reduce the risk perceptions on the part of their perspective customers (e.g., Gemuenden, 1985). The FSPs should provide offer details in the cover letter, use comprehensive brochures to convey additional information, and provide a “request information” response option to accommodate the further information demands of the direct mail recipients.

- Use personalization as a differentiating factor. Most organizations that send out personally addressed direct mailings employ certain personalization techniques by default (e.g., personal salutary address). These personalization features have become commonplace and, thus, can hardly continue to serve as distinguishing design factors. However, personalization of the additional parts of the mailing such as supplements is still not commonly employed; as shown in Table 2, only 64 out of 479 campaigns containing a supplement are personalized to the recipient. Under these circumstances, personalization can serve as an effective differentiator.

- Take little steps when approaching prospects. Another interesting parallel is that neither the NPO nor the FSP campaigns should be too intrusive. In prospecting, charitable direct mail solicitations should aim at one-time donations rather than formal and enduring new members (see Table 4). Similarly, the FSPs should be cautious about immediately requesting that the prospective customer sign a contract in the direct mail solicitations. Likewise, mentioning restrictive terms and conditions up front in the letter should be avoided. The focus of financial mailings should rather be on initiating a promising customer acquisition process by providing sufficient information in the letter and supplement and by offering a request for further information as a response option in case the prospect is interested (see Table 4). As shown in Table 2, common industry practice is not consistent with this finding: out of 225 FSP campaigns containing a response device, 87 campaigns aim to have an application form filled out right away, whereas only 31 campaigns offer a request for additional information as a response option. Our general recommendation, however, is that the solicitations that aim at substantial immediate commitments by prospects should only be considered if the final payoff of the completed applications or memberships overcompensates for the lower KRs.

9. Future research

The data-driven constraints in our study indicate areas for future research. First, our study is the first to investigate direct mail characteristics in Germany. While Germany is one of the largest economies in the world, it has to be taken into account that the households in Europe receive a much smaller number of direct mail solicitations than those in the United States (Hesse, Krafft, & Peters, 2007). One could argue that consumers’ preferences and attitudes towards direct mail advertising vary between countries. Accordingly, the effects of specific mailing characteristics on the various measures such as opening, keeping, and response rates can differ. Hence, there is a need for an international study that covers multiple countries, preferably from different continents including the Americas and Asia. Second, the actual response rates were not available to us. While we argue that the ORs and the KRs of the campaigns as intermediate measures reflect the effectiveness of envelope and direct mail design characteristics more accurately, the inclusion of actual response rates would have been desirable to put our results into perspective. Third, while our study comprises a wide range of design variables, it is possible that there are additional design features that also influence campaign performance. Similarly, different operationalizations of our variables could produce different results. For example, letter length was measured by a dummy variable indicating whether the letter is longer than 1 page. We demonstrate positive effects of letter length on the KRs. More fine-grained measures such as the number of lines or words would have permitted us to test for linearity such as inverted U-shape effects. Fourth, individual-level data are not available to us because GfK only provided us with aggregate data at the campaign level. A disaggregate analysis could provide deeper insights into the effectiveness of mailing characteristics across individuals. Customer characteristics as well as unobservable factors, such as attitudes or preferences, can be included in this type of estimation framework and could result in an even better understanding of consumer behavior in the context of direct mail advertising.

Acknowledgments

The authors thank the Editor, the Area Editor and the anonymous referees for their support, guidance, and encouragement. Ashwin Aravindakshan, Norris I. Bruce, Bas Donkers, Edward Fox, Edward C. Malthouse, Prasad A. Naik, Kalyan Raman, the participants of the 2008 DDV scientific congress, 2009 Marketing Science at University of Michigan, 2009 Joint Statistical Meeting in Washington DC, 2009 PhD Seminar Rothenberge, 2009 SALTY Conference at Kiel University, and 2010 DMEF Conference at San Francisco provided valuable feedback on earlier versions. The authors thank Deutsche Post DHL, the SVI Siegfried Vögele Institute, and GfK for their support with financial resources, insights and data. The authors thank Christina Gnacke and Brigitte Jansen for their initial support in collecting and preparing the raw data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2012.07.003.

References


User-generated versus designer-generated products: A performance assessment at Muji

Hidehiko Nishikawa a,1, Martin Schreier b,⁎, Susumu Ogawa c,2

a Hosei University, Faculty of Business Administration, 2-17-1 Fujimi, Chiyoda-ku, Tokyo, 102-8160 Japan
b WU Vienna University of Economics and Business, Marketing Department, Augasse 2-6, Vienna, 1090 Austria
c Kobe University, School of Business Administration, 2-1 Rokkodai-cho Nada-ku, Kobe Hyogo, 657-8501 Japan

A R T I C L E   I N F O

Article history:
First received in 6, June 2012 and was under review for 5 months
Available online 17 October 2012

Area Editor: John H. Roberts

Keywords:
User design
User-generated products
User innovation
Co-creation
Customer integration
Crowdsourcing
Market performance
New product development
Idea generation

A B S T R A C T

In recent years, more and more consumer goods firms have started to tap into the creative potential of their user communities to fuel their new product development pipelines. Although many have hailed this paradigm shift as a highly promising development for firms, hardly any research has systematically compared the actual market performance of user-generated products with designer-generated ones. We fill this void by presenting a unique data set gathered from the Japanese consumer goods brand Muji, which has drawn on both sources of ideas in parallel in recent years. We demonstrate that user-generated products, which are found to generally contain higher novelty, outperformed their designer-generated counterparts on key market performance metrics. Specifically, in the first year after introduction, sales revenues from user-generated products were three times higher and gross margins were four timesgreater than those of designer-generated products. These effects also increased over time: after three years, the aggregate sales revenues of user-generated products were, on average, 1.25 billion yen (approximately 16 million dollars) higher, or five times greater, than the sales of designer-generated products. The corresponding average margin was an impressive 619 million yen (approximately 8 million dollars) higher, or six times greater, than the margin for designer-generated products. Finally, user-generated products were more likely to survive the three-year observation period than designer-generated products (i.e., were still on the market three years after introduction). These findings clearly favor the paradigm shift identified in marketing research and appeal to managers considering the integration of user ideas into the process of new product development. We discuss our study’s limitations and identify important avenues for future research.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

For decades (if not for much longer), professional marketers, designers, and engineers rather than consumers or users have been the dominant agents in the development of new consumer products. Although it has always been imperative to listen to customers to discern emerging market trends and unmet consumer needs, design professionals employed by firms (or their subcontractors) are usually responsible for translating the resulting opportunities into new product ideas and market offerings (Cooper, 2001; Crawford & Di Benedetto, 2000; Ulrich & Eppinger, 1995; Urban & Hauser, 1993). Until very recently, firms did not consider user involvement in the process of generating ideas for new product development (NPD) to be particularly promising. For example, as Schulze and Hoegl (2008, p. 1744) note, “relying on the method of asking buyers to describe potential future products, big leaps to novel product ideas are generally not likely.” One of the key assumptions underlying this dominant idea generation paradigm is that a firm’s professionals, unlike customers or users, have the requisite expertise to invent new and useful product concepts (e.g., Amabile, 1998; Weisberg, 1993). A firm’s professionals “have acquired skills and capabilities that allow them to perform most design tasks more effectively and at a higher level of quality” (Ulrich, 2007, Chapter 3, 5ff) and they “often have a significant advantage [...] over consumers, in terms of their knowledge, training, and experience” (Moreau & Herd, 2010, p. 807).

However, research on the sources of innovation (Von Hippel, 1988) has long challenged this fundamental assumption and the related, widely held view that users are of little value to firms in providing ideas to solve their unmet consumer problems (as opposed to merely pointing out such problems to firms in the first place). In particular, this line of research has found that many major and minor innovations across various industries were originally developed by users rather than professionals in firms (cf. Von Hippel, 2005). In their seminal work,

http://dx.doi.org/10.1016/j.ijresmar.2012.09.002
Lilien, Morrison, Searls, Sonnack, and von Hippel (2002) provide further evidence that active user involvement in idea generation might benefit a firm’s NPD efforts, at least in industrial markets. Specifically, Lilien et al. (2002) compare the potential value of several new industrial product concepts jointly developed by 3 M personnel and selected lead users to those developed by more conventional means (i.e., internal developers only). They found that the former outperformed the latter on key innovation indicators that were assessed by 3 M managers (e.g., the product concept’s novelty compared to the competition or its potential to create an entire new product family). Furthermore, they found that the sales forecasts for concepts developed by lead users were eight times higher than those of internally developed ideas.

One might plausibly argue that working with knowledgeable industrial users is one thing, while working with potential end consumers is another. However, in various consumer goods domains as well, users are frequently found to innovate for themselves, and many of their innovations are commercially attractive (Franke, von Hippel, & Schreier, 2006). For example, a recent survey of a representative sample of UK consumers revealed that six percent, or nearly three million consumers, innovated in the domain of household products. Similar national user innovation statistics have been reported for other countries, including the US (5%, or almost 12 million consumers) and Japan (4%, or almost 4 million consumers) (von Hippel, Ogawa, & de Jong, 2011).

Thus, user innovation has come of age as many consumer goods firms are challenging the traditional paradigm and experimenting with new ways to more actively integrate users into the idea generation process. In extreme cases, firms like Threadless no longer employ designers but rely exclusively on their user communities to generate new products (Ogawa & Piller, 2006). Many established firms and brands, such as LEGO, Dell, and Muji, have followed suit and now complement their in-house efforts with public idea contests known as “crowdsourcing” initiatives, where idea generation is outsourced to a potentially large, unknown population (the crowd) in the form of an open call (Bayus, 2010; Howe, 2008; Surowiecki, 2004). Thus, crowdsourcing relies on a self-selection process among users who are willing and able to respond to widely broadcast idea generation competitions (Jeppesen & Lakhani, 2010; Piller & Walcher, 2006; Poetz & Schreier, 2012).

Conceptually, the key insight is that some (but, of course, not all) ideas generated by these self-selected users might outperform the ones generated by firm designers (Poetz & Schreier, 2012; Schreier, Fuchs, & Dahl, 2012). Users may have a competitive edge in idea generation over designers through their experience as consumers. Leading-edge users, in particular, may have tried to solve consumption problems themselves, and their ideas may be commercially attractive because they foreshadow what other consumers will demand in the future (cf. Von Hippel, 2005). The respective user population that might be activated via crowdsourcing is also naturally much larger and more diverse compared to the team of designers employed by a given firm; a firm’s user community may comprise thousands of talented users from highly diverse backgrounds. It follows that the generation of more (and more diverse) ideas increases the odds of generating a few truly exceptional ones (e.g., Gross, 1972; Schreier et al., 2012; Surowiecki, 2004; Terwiesch & Ulrich, 2009).

Empirically, researchers have recently begun to address important questions like what motivates users to participate (e.g., Hertel, Niedner, & Herrmann, 2003), how idea contests should be organized (e.g., Boudreau, Lactera, & Lakhani, 2011), how user ideas can be screened effectively (e.g., Toubia & Flores, 2007), or how consumers perceive user-driven firms (e.g., Fuchs & Schreier, 2011). In a recent study, for example, Schreier et al. (2012) find that consumers associate user-driven firms with higher innovation abilities, that is, they perceive such firms as being better able to generate innovative new products. In a series of experimental studies, they further find that this innovation inference prompts consumers to demand one and the same set of products more strongly. This finding is especially interesting because objective product properties were kept constant between experimental conditions.

From an NPD perspective, however, the important question is not if the design mode affects consumer perceptions of innovation ability, but rather if it affects the objective quality of new product ideas actually generated. This question has recently been addressed by Poetz and Schreier (2012), who compared the quality of ideas for new baby products generated by a firm’s designers to those generated by users in a public idea contest. Ideas from both sources were judged by company executives (who were blind to the source of ideas), and the key finding was that user ideas were characterized by higher novelty and higher customer benefit, but also by lower feasibility.

Although these findings are promising, it has yet to be explored whether the best of these user ideas, if realized, would eventually perform better on the market. There are several reasons that make this extrapolation non-obvious. First and foremost, it should be noted that in most cases firms still have to translate user ideas into marketable products; a process that demands much effort and involves many decisions that may or may not lead to ultimate market success. More specifically, lower feasibility scores of user ideas point to the possibility that the related promise might never be realized. Second, and on a related note, it could be that high innovativeness and low feasibility eventually boost development and production costs and thereby reduce any potential margin. Third, it is always an empirical question how well managers’ perceptions of ideas match consumer reactions to new products once they reach the shelves. Even if this match is strikingly high (i.e., no judgment and translation errors are involved), it remains unclear how a certain increment in innovativeness (e.g., 10%) will translate into incremental market performance (e.g., X% more sales). Finally, it is worth noting that firms can only market a very limited number of new products at a given point in time. As such, what matters is not the entire distribution and the average idea quality, but rather the extreme values: the very best ideas available (e.g., Fleming, 2007; Singh & Fleming, 2010). It remains unclear whether the very best ideas from both sources would differ to such an extent that one could observe managerially meaningful differences in the ultimate consumer reactions.

Against this backdrop, it is surprising that a systematic, empirical market performance assessment of user- vs. designer-generated consumer products is still missing in the literature, despite the broad and growing interest in this topical phenomenon. This is most likely due to the difficulty of obtaining reliable quantitative market data from real-world practice. Such research is important because the market performance of user-driven initiatives will guide marketers in deciding whether to follow the paradigm shift and involve users in generating new product ideas. Drawing on a unique data set gathered from the Japanese consumer goods brand Muji, we address this important gap in the literature. In short, our study reveals that user-generated products systematically outperform designer-generated ones in terms of important outcome variables, including actual sales revenues. Most importantly, we see our study as “initial evidence” in favor of user-generated products and as a contribution to a more critical reflection on the dominant NPD paradigm. As such, this paper makes a second contribution in its conceptual discussion of the generalizability of our findings.

2. Study method

2.1. New product development at Muji

Muji began experimenting with idea contests ten years ago; thus, it is a forerunner among firms that market user-generated products on a broad scale. Muji has also recently marketed a number of user-generated products alongside products generated by the firm’s designers, which allows us to empirically assess the performance of the two approaches. In addition to granting us access to sensitive
data (e.g., sales data, gross margins), the firm provided us with substantial support to complement and validate performance data (e.g., interviews with executives, coding of ideas).

Muji is an established Japanese manufacturer and retailer brand of a broad range of consumer goods of Ryohin Keikaku Co., Ltd, with a particular focus on interior and household products. Muji’s products are sold in almost 500 Muji stores in 22 countries including Japan, China, the US, the UK, France, Italy, and Germany. In 2010, Muji generated approximately 170 billion yen in sales revenue (approximately 2.2 billion dollars), producing net profits of almost 8 billion yen (approximately 100 million dollars). Over the last few decades, Muji has become known for developing and marketing innovative consumer solutions, which is reflected in their receipt of more than 90 design and innovation prizes (e.g., two international Design Oscars awarded by IF International Forum Design, two Good Design Award Gold awards from the Japan Institute of Design Promotion, one Michael E. Porter Prize for innovation). Their ability to generate successful innovations sets a high standard of performance for user-generated products.

For the purposes of our research, we focus on Muji’s furniture division. By focusing on one product type, we minimize the potential of confounding effects that arise from comparing different product categories. The furniture division is strategically important to Muji because it accounts for nearly 20 percent of their total sales. We also selected Muji’s furniture division because it has experimented more with user-generated products in recent years than other divisions; in other words, data for the furniture division contained the largest number of observations for user-generated products. One might argue that this method of sample selection may bias our results because the furniture division began to rely on user input due to the underperformance of designers, thereby lowering the performance standard that user-generated products must exceed. In our interviews at Muji, however, we found no evidence for such selection bias.

In 2002, Muji launched the first user-driven piece of furniture, the “Floor Sofa,” which is a large cushion that sits on the floor and adjusts to the body. Because of the novelty and experimental nature of this project, Muji collected binding customer pre-orders before investing in final product development to minimize the risk of failure. The Floor Sofa was a great success. Within the first year, it generated almost 500 million yen in sales, a record that was not exceeded by any of Muji’s designer-generated products until the beginning of our observation period (2005). Prior to 2005, the firm carried out another crowdsourcing experiment that led to the production of a new piece of furniture in 2003 (a small, movable wall shelf) that generated first-year sales of 71 million yen (ranking it sixth among the 15 new furniture products introduced in 2003).

Based on the success of these products, the firm decided to use crowdsourcing on a more systematic basis. During our observation period (February 2005 to July 2009), 43 new products were developed, produced, and introduced to the Japanese market. Thirty-seven products were designer-generated, and six were user-generated. We excluded the two previous user-generated products from our analysis because they differed from the other products in terms of the idea selection process (binding customer pre-orders compared to selection by the firm). However, the results do not change if the observation period is extended to include these two early user-generated products, as well as the respective designer-generated products (results available from the authors upon request). In what follows, we describe the development of designer- and user-generated products in greater detail.

2.1.1. Designer-generated new products

Muji’s conventional approach to NPD adheres closely to the recommendations made in the traditional literature on marketing and innovation. The active agents in idea generation are the firm’s designers, engineers, and marketers, who work in teams organized around specific, new product development projects. Based on established market research (e.g., market trend studies, focus groups with target consumers), the team identifies and selects important consumption problems to be addressed by a new product (referred to as new product “themes”). In one product development project, for example, the theme was limited storage capacity in consumers’ bedrooms due to small room sizes. Based on insight from market research, members of the project team generate new product ideas (e.g., a flat-panel shelf matching the firm’s line of beds). The best ideas enter the next phase of product development, where designers propose more detailed design concepts and product specifications based on the product’s intended functionality (e.g., a specific design of a flat-panel shelf that is as thin as possible yet large enough to guarantee proper use and stability). The team then decides what concept to move to the next stage. Finally, a new product is introduced to the market. The key characteristic for our study is that it is the internal employees who generate new product ideas aimed to address the selected consumer problem (i.e., the new product theme).

2.1.2. User-generated products

In a second development track, Muji invites users to generate ideas for new products. Anyone who registers on their website can participate. Similar to the internal development of designer-generated products, idea generation follows a specific theme for which solutions can be proposed. Users can also strongly influence the selection of the theme by participating in online discussion boards (as is the case in the designer-generated track). The theme “Suwaru Seikatsu” (“Sit Down Life”), which aimed to generate ideas for better and more comfortable sitting solutions in consumers’ homes, was one of the first themes to be successfully outsourced to users. The theme was broadcast to users, who responded with many new product ideas, including the eventual winner: the Floor Sofa. After the first stage of idea generation, users can also comment, vote, and improve on each other’s ideas. The best idea is adopted by a Muji project team, which defines the product characteristics in more detail (e.g., external fabrics, cushion filling), seeks further feedback from the user community, and finally moves the new product to the full production stage. Thus, Muji employees are also heavily involved in the development process for new products in this second track. In contrast to the first track, however, new product ideas are generated by users rather than the firm’s designers.

Adopted user ideas in our observation period included the following: a sofa-bed (a sofa that can easily be changed to a bed and vice versa); a wooden cupboard with a movable wagon to increase flexibility of use; a small sofa chair with slim arms to fit in small rooms; a highly versatile stacking shelf; a bed with a thin, yet robust, wooden frame to provide more storage space under the bed; and a sofa with arm rests that are comfortable for both sitting (arms) and lying down (head and legs).

A series of in-depth interviews with senior managers and project members revealed that the two design tracks did not differ substantially on other important dimensions. In particular, projects centered around designer-generated vs. user-generated products did not differ in terms of the human resources invested by the firm (e.g., the quantity and quality of project staff). Furthermore, it is unlikely that employee motivation differed substantially between the two development processes because the firm’s employees usually work on several projects at once (including designer- and user-generated products), and both individual and team incentives are identical across projects. Finally, the firm’s idea and concept selection criteria are identical in both tracks; a minimum threshold of expected sales and gross margins must be met before a project moves to full production. Interviews with managers further refuted the argument that adopted user-generated ideas “survived a tougher screening” process. Therefore, it is unlikely that any differences in performance between designer- and user-generated products are attributable to these alternative factors.

Nonetheless, the number of ideas generated differed systematically between the two processes, which is consistent with the theoretical
work on user involvement and crowdsourcing (Poetz & Schreier, 2012; Schreier et al., 2012). The average estimated number of ideas per theme was approximately ten in the designer condition (this number did not vary much across projects because teams are expected to come up with at least ten good ideas before a choice is made), whereas the lowest number of ideas per theme was over 400 in the user condition (the observed maximum was over 2000). Any random user idea, therefore, has clearly lower odds of being selected, ceteris paribus. However, this does not bias our results. If both distributions empirically ended at the same right-hand point and these top ideas were selected, there should still be no difference in market performance, as there were no differences in the selection criteria employed. If, however, extreme values of the user distribution stretched farther to the right (e.g., because having more ideas increases the likelihood of generating some extremely high values; e.g., Gross, 1972; Terwiesch & Ulrich, 2009), the best user-generated ideas may very well perform better. Such a scenario is consistent with the theoretical arguments that favor user involvement.

2.2. Measures

2.2.1. Independent variable

The independent variable is the source of idea generation: professional designers vs. users (i.e., user-generated vs. designer-generated products).

2.2.2. Dependent variables

Our key outcome variable is the products’ actual market performance, measured as aggregate unit sales generated in the first year after market introduction. In addition, we capture profitability by measuring the products’ monetary value of sales and gross margins. Because the firm does not allocate fixed costs to individual products, we cannot assess the first-year profits of products, but only their gross margins. Data from the first year that a product is on the market form the most important basis for a product’s performance assessment by Muji because furniture products are characterized by short product life cycles (i.e., the first year usually produces the highest sales). To check the robustness of our results, we also gathered aggregate sales data three years after a product was introduced to the market (the average product life cycle of Muji products is approximately three years), as well as product survival information (i.e., information about which products survived our observation period and were still on the market three years after introduction).

In addition to performance data, we coded all new products (in random order) based on important innovation indicators to gain insight into the nature of the underlying product ideas (e.g., Lilien et al., 2002). First, we asked the general manager of the furniture division and the person in charge of each new product development project to rate the products’ novelty relative to competing products and its novelty based on customer needs addressed (measured on a 10-point scale where 1 = low and 10 = high). Because both measures of novelty were highly correlated for both raters (rRater 1 = .52, p < .001, rRater 2 = .84, p < .001), we averaged these measures to create a novelty index. The high correlation between raters (rBetween raters = .51, p < .001) suggests that novelty is captured by this index in a valid way (e.g., Franke et al., 2006). We also captured the products’ strategic impact on the firm’s product portfolio, which we measured using two items: the strategic importance of the product to the business unit’s future and the potential of the product to generate a new product line (measured on a 10-point scale where 1 = low and 10 = high). Because both measures were highly correlated for both raters (rRater 1 = .50, p < .01, rRater 2 = .43, p < .01), we averaged them to create a strategic impact index. The high correlation between raters (rBetween raters = .47, p < .01) suggests that this variable captures a product’s strategic impact in a valid way.

2.2.3. Control variables

We considered the products’ price and their year of market introduction as control variables. We re-ran our main analysis (reported below) and controlled for these variables in OLS regressions. As we find that neither price nor inter-year variations in sales, etc. account for our findings (all reported effects remain significant), we do not report the resulting statistics due to space constraints (results are available upon request).

3. Findings

To compare the market performance of user-generated and designer-generated products, we compared the means of market performance measures for both types of products and tested their statistical significance using t-tests. Given the small sample size of the user design condition (n = 6), we also performed Mann–Whitney U-tests to assess the robustness of our results. The results indicate that both tests lead to the same conclusions (see Table 1A).

3.1. Unit sales

In the first year after introduction, we find that user-generated products sell roughly twice as much as designer-generated products (MUser = 30,182.33, MDesigner = 14,191.76, pبيضercise < .05). Similarly, we find that the distribution of the products’ first-year unit sales is significantly different between the two conditions (pmW-U < .05). Five out of six user-generated products are represented in the top third of the best sellers among all 43 products (ranked 2, 4, 7, 10 and 11), and only one user-generated product had lower unit sales than the average designer-generated product (see Fig. 1). The results are even more pronounced using aggregate sales data across three years: units of user-generated products sold three times more frequently than designer-generated ones (pR-test < .12, pM-W-U < .05).

3.2. Sales

First-year sales revenues of user-generated products are more than three times higher than those of designer-generated products (MUser = 500.97 million yen, MDesigner = 141.04 million yen, p bitwise < .05). Similarly, the distribution of sales revenues differs significantly between the two types of products (pmW-U = .001). Four of the top six products, in terms of sales, are user-generated (ranked 1, 2, 3, 6), and the two remaining user-generated products generated higher sales revenues than the average designer-generated product (greater than 147.80 million yen, see Fig. 2). Aggregate sales data across three years suggests that the size of this effect increases over time: three years after introduction, user-generated products generated an average of 1250 million yen (approximately 16 million dollars) more in sales than designer-generated products, which is equal to over five times the sales of designer-generated products (pR-test < .05, pM-W-U = .001).

3.3. Gross margin

User-generated products yielded gross margins (MUser = 240.47 million yen) that were approximately four times higher than designer-generated products (MDesigner = 56.22 million yen, pR-test < .05). The distribution of gross margins is also significantly different between the two conditions (pmW-U = .001). In particular, four of the top five products, in terms of gross margins, are user-generated products (ranked 1, 2, 3 and 5) and the remaining two user-generated products yielded higher gross margins than the average designer-generated product (greater than 67.90 million yen; see Fig. 3). The results are more pronounced if we use three years of sales data: the average margin of user-generated products is 619 million yen (approximately 8 million dollars) higher, or six times greater, than designer-generated products.
Because performance should be related to survival, we additionally analyzed how many designer- and user-generated products survived our observation period (three years after introduction). Products that survived performed substantially better than those which were discontinued (e.g., first-year unit sales were approximately 60% higher for products that survived the third year). Indeed, only 17 out of 37 designer-generated products were still on the shelves after three years. In contrast, five out of six user-generated products were still on the market \((\text{Pt-test} < 0.01; \text{Pm-W-U} = 0.001)\). User-generated products, therefore, have a substantially higher survival rate at Muji.

### 3.4. Novelty and strategic impact

The results with regard to the nature of each new product provide corroborating evidence for the findings reported above. In particular, we find that user-generated products are characterized by substantially higher novelty compared to designer-generated products \((\text{MUser} = 7.29, \text{MDesigner} = 4.78, \text{Pt-test} < 0.01, \text{Pm-W-U} < 0.001)\). Four of the top five products, in terms of novelty, are user-generated (ranked 1, 2, 3 and 4) and the remaining two user-generated products are rated among the top 50% of the sample on novelty and consistently exhibit higher novelty than the average designer-generated product (greater than 5.90, see Fig. 4). The results for the products’ strategic impact are similar: user-generated products have a significantly more pronounced strategic impact on the firm’s future product portfolio than designer-generated ones \((\text{MUser} = 6.08, \text{MDesigner} = 4.25, \text{Pt-test} < 0.01, \text{Pm-W-U} < 0.001)\). Three of the top five products, in terms of strategic impact, are user-generated (ranked 1, 3 and 4), and the “worst” user-generated product has a strategic impact score (4.00) close to the average score of designer-generated products (see Fig. 5).

### 3.5. Sensitivity analysis

Although we only analyzed products within a single category (furniture), the sample consists of different products (e.g., sofa, table, chair, bed, shelf). To assess whether different product types within the furniture category might partially account for the findings reported above (beyond price considerations), we re-analyzed the data using a more homogeneous subsample of products. We matched the six user-generated products with similar designer-generated ones \((n = 15; \text{sofa, bed, shelf})\). As Table 1B shows, the results remain robust; user-generated products outperform designer-generated ones on all dimensions.
In a second robustness check, we aimed to broaden the scope of our analysis. In particular, we sought market performance data from another category, namely health and beauty products. This category has brought about four user-generated products in recent years (in the area of natural and homeopathic remedies and skin care) where the respective product launch was far enough in the past in order to track aggregate performance data across three years. Table 2 shows that user-generated products once again demonstrate better market performance. Although we identified several other Muji categories with user-generated products (e.g., kitchenware, stationery, gardening products, and clothing), user-generated products in these categories were either too low in number (e.g., one or two products) and/or market introduction had occurred too recently (e.g., in 2010).

4. Discussion

There is broad and growing interest in understanding the topical phenomenon of more actively involving users in firms’ idea generation processes. Yet and despite conceptual advancements (Poetz & Schreier, 2012; Schreier et al., 2012) it is surprising that an empirical market performance assessment of user-driven products (i.e., products that are based on user- vs. designer-generated ideas) is still missing in the literature. This is most likely due to the difficulty of obtaining reliable market data from real-world practice. As such, it remains unclear whether it will pay off for firms to depart from the classic NPD paradigm anchored in the idea that firm professionals – and not users – should generate the ideas for new products to be marketed to broader customer segments. By drawing on a unique data set gathered from the Japanese consumer goods brand Muji, which has drawn on both sources of ideas in parallel in recent years, we provide initial research to fill this gap in the literature. We demonstrate that user-generated products systematically and substantially outperform their designer-generated counterparts in terms of key market performance metrics.

Of course, we do not see our study as the “final battle” that would once and for all relegate the “loser” to the history books. Rather, we see our study as initial evidence and as a contribution to a more critical reflection on the dominant NPD paradigm. In a nutshell, our findings suggest that it might pay for firms to draw on user ideas in parallel to their established in-house efforts — not unlike what Muji seems to have identified as their best practice. Furthermore, users were only the providers of ideas; the overall NPD process entails many more stages, and it goes without saying that decisive in-house efforts and capabilities are needed to convert any promising idea into a successful new product. Bearing this in mind, we now discuss the generalizability of our findings and identify promising directions for future research.

First, it is worth noting that the majority of winning user-designers in our sample were highly involved in the category, had...
enduring past exposure to the problems for which ideas were sought, had relatively high levels of technical expertise, and thus generally matched the conceptual description of lead users (cf. Von Hippel, 2005). Clearly, if firms are unable to activate this type of user, it is unlikely that the effects found in this study can be replicated. The notion of users’ leading-edge status is also conceptually important. Different types of users might account for the discrepancy between predictions made in the classical marketing and NPD literatures, on the one hand, and research on user innovation, on the other hand. Whereas the former might have “average” users in mind (which would justify skepticism about user involvement because these users lack both technical and usage expertise), the latter typically constructs their arguments around leading-edge users. In other words, conventional marketing is concerned with the “average user” and research on user innovation targets only users with leading-edge status who represent “extreme values” on the right-hand side of the distribution of users.

The consequence of involving the “right” vs. “wrong” users might loom even larger for more complex products than the ones studied in this paper (Poetz & Schreier, 2012; Schreier et al., 2012). For complex products (e.g., cars), the expertise required for successful ideation might be too high, thereby preventing the vast majority of average users from contributing meaningful ideas. On a related note, all of the furniture themes studied were based on a clear “usage problem” (e.g., improving the sitting experience in the living room). We argue that such problems are well-suited for users because users have a competitive edge over designers in terms of the needs-based information necessary to solve the problem effectively. If, in contrast, the theme was based on technical details regarding technologies or materials, company designers would be more likely than users to have the skills and expertise to solve these problems. These speculations offer promising avenues for future research.

Furthermore, the openness of contests may favor user-based ideas. The user population that is accessed through crowdsourcing is naturally much larger and more diverse compared with a team of designers employed by a given firm. It follows that, if more (and more diverse) ideas are generated, the odds of generating a few truly exceptional ideas will increase (e.g., Gross, 1972; Schreier et al., 2012; Surowiecki, 2004; Terwiesch & Ulrich, 2009). Our data from Muji support this prediction: out of the many ideas submitted by users (ranging from 400 to more than 2000 per theme), the best user ideas outperformed those generated by designers. Yet the relationship between the openness of contests, how the contest announcement is targeted, the activation of effective self-selection mechanisms, and the final outcome is not well understood; thus, it constitutes fruitful ground for future research.

Interviews with managers at Muji further revealed that they have succeeded in building the specific capability to interact with users effectively. Nonetheless, each crowdsourcing initiative took substantial time to implement, is generally less controllable and manageable in the early stages, and it cannot easily be implemented on a large scale.
Note that we also analyzed survival in this category. Importantly, we again only 21 out of 49 designer-generated products were still on the shelves after three years. Such efforts might pay off.

Finally, it is worth noting that Muji did not actively market user-driven products to the general public. We, therefore, should examine this psychological effect on performance.

Acknowledgments

The authors thank Muji, in particular Satoshi Shimizu, Tsunemi Kawana, Tetsu Takahashi, Tomisaburou Hagiwara, Yukie Saito and Ikuo Tuskari, for their support in realizing this project. They further thank Pradeep K. Chintagunta, Berend Wierenga, and research seminar participants at the Universitat Pompeu Fabra Barcelona, Università di Bologna, and VU University Amsterdam for their feedback on earlier drafts of the manuscript as well as Marnik G. Dekimpe, the AE, and the IJRM reviewers for their helpful suggestions and guidance during the review process. Financial support is greatly acknowledged from JSPS KAKENHI (Grant Numbers 22330130, 24530535) and the Danish Strategic Research Council (Grant Number 09-067157).

References


Surowiecki, James (2004). The wisdom of crowds — Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York: Doubleday.


Table 2

| (A) Analysis of the full sample | Designer-generated products (n=49) | User-generated products (n=4) | Significance |
| M | M | M | M |
| Unit sales (after 1st year) | 45,527.76 | 74,787.50 | p<0.01; r<0.1; t<0.05 |
| Unit sales (after 3rd year) | 106,452.61 | 170,995.25 | p<0.01; r<0.1; t<0.05 |
| Sales in yen (after 1st year) | 13.00 million | 69.10 million | p<0.01; r<0.1; t<0.05 |
| Sales in yen (after 3rd year) | 28.63 million | 162.64 million | p<0.01; r<0.1; t<0.05 |
| Gross margin in yen (after 1st year) | 4.38 million | 23.50 million | p<0.01; r<0.1; t<0.05 |
| Gross margin in yen (after 3rd year) | 9.97 million | 57.50 million | p<0.01; r<0.1; t<0.05 |

| (B) Sensitivity analysis (product-type matching of user- and designer-generated products) | Designer-generated products (n=12) | User-generated products (n=4) | Significance |
| M | M | M | M |
| Unit sales (after 1st year) | 38,428.50 | 74,787.50 | p<0.01; r<0.1; t<0.05 |
| Unit sales (after 3rd year) | 74,878.50 | 170,995.25 | p<0.01; r<0.1; t<0.05 |
| Sales in yen (after 1st year) | 7.20 million | 69.10 million | p<0.01; r<0.1; t<0.05 |
| Sales in yen (after 3rd year) | 12.14 million | 162.64 million | p<0.01; r<0.1; t<0.05 |
| Gross margin in yen (after 1st year) | 2.05 million | 23.50 million | p<0.01; r<0.1; t<0.05 |
| Gross margin in yen (after 3rd year) | 3.99 million | 57.50 million | p<0.01; r<0.1; t<0.05 |

Note that we also analyzed survival in this category. Importantly, we again find that only 21 out of 49 designer-generated products were still on the shelves after three years. At the same time, however, all four user-generated products were still on the market [Phi-square difference test < 0.05]. User-generated products thus demonstrate a substantially higher survival rate. (Products that survived the three-year observation period also performed substantially better than those that were discontinued, which confirms that lower-performing products were eliminated; e.g., first-year unit sales were approximately 140% higher for products that survived the third year.)
Disentangling the market value of customer satisfaction: Evidence from market reaction to the unanticipated component of ACSI announcements☆

Vladimir Ivanov a,1, Kissan Joseph b,*, M. Babajide Wintoki b,2

a Office of Economic Analysis, U.S. Securities and Exchange Commission, 100 F Street, N.E. Washington DC 20549, United States
b School of Business, University of Kansas, 1300 Sunnyside Avenue, Lawrence, KS 66045-7585, United States

ABSTRACT

There is a rich literature that examines the impact of customer satisfaction on market value. Surprisingly, the short-run market impact of customer satisfaction has been found to be either insignificant or limited in scope. To address this shortcoming, we introduce the notion that investors form expectations about customer satisfaction and respond only to deviations from these expectations (i.e., “surprises”). We consider two “expectations” models: a naïve model that utilizes last year’s scores and a model that includes firm characteristics and marketing investments to proxy for the prior allocation of resources devoted to improving customer satisfaction. In our empirical work, we find that the market does indeed respond in the short-run to surprises in customer satisfaction, with more pronounced effects for our second expectations model. Overall, our research offers two distinct contributions. First, it refines the current conceptualization of customer satisfaction by explicitly introducing the notion of investor expectations. Second, we employ this refined conceptualization to unequivocally demonstrate the short-run impact of investments in customer satisfaction.

1. Introduction

1.1. Motivation

There is a rich literature that examines the impact of customer satisfaction on important outcomes in the financial market. As is worthy of a major field of inquiry, the existing works have examined various elements of this relationship. Some studies, such as those of Fornell, Mithas, Morgeson, and Krishnan (2006), Ittner and Larcker (1998), and Ittner, Larcker, and Taylor (2009), investigate the immediate, short-run market impact of customer satisfaction announcements via event studies. Surprisingly, these researchers report effects that are either insignificant or limited in scope. Specifically, across two separate event studies, Fornell et al. (2006) and Ittner and Larcker (1998) find no short-run impact of customer satisfaction announcements on stock prices. More recently, Ittner et al. (2009) also employ the event-study methodology. They find that investors respond to customer satisfaction announcements but only if there are large positive changes over the levels of previous years.

Other studies investigate the long-run market impact of customer satisfaction. Within this stream, Anderson, Fornell, and Mazvancheryl (2004), for example, find a positive association between customer satisfaction and shareholder value as measured by Tobin’s Q. In effect, they demonstrate that customer satisfaction is value-relevant in the long-run: higher levels of customer satisfaction predict higher values of Tobin’s Q, even after accounting for fixed, random, and unobservable factors. Other studies have also focused on the mispricing of customer satisfaction by the financial markets. Within this stream, Fornell et al. (2006) show that customer satisfaction helps predict long-run changes in equity prices. Specifically, portfolios composed of firms with high customer satisfaction scores outperform the market even after accounting for trading costs. This particular finding pertaining to the ability of portfolios with high customer satisfaction firms to obtain abnormal returns has subsequently been subject to additional scrutiny in a series of articles. In particular, Jacobson and Mizik (2009a) conclude that the mispricing reported in the existing literature is due to a small group of satisfaction leaders in the computer and internet sectors. This conclusion is further echoed by Ittner et al. (2009). Similarly, O’Sullivan, Mark, and Hutchinson (2009) correct for three interlinking issues: risk adjustment, abnormal risk adjustment, and portfolio aggregation, and find no evidence of mispricing.
However, in their commentary, Fornell, Mithas, and Morgeson (2009) present new analysis that updates the findings of Fornell et al. (2006). They report that the above-market returns reported in Fornell et al. (2006) persist; moreover, they are both economically and statistically significant.3

Clearly, the marketing scholars and practitioners in all of these investigations are fundamentally interested in better understanding the market value of customer satisfaction. Indeed, the chain of effects leading from customer satisfaction to favorable customer behaviors, followed by enhanced firm performance and culminating in a positive stock market response has received substantial attention in the literature. As summarized by Anderson and Mansi (2009), the relationship between customer satisfaction and a host of favorable customer outcomes, such as acquisition costs, retention, word-of-mouth, willingness to pay, usage, cross-selling opportunities, reduced complaints, and lower payment defaults, now appears to be widely accepted and credibly established (Anderson, Fornell, & Lehmann, 1994; Bearden & Teel, 1983; Bolton, 1998; Bolton & Drew, 1991; Bolton & Lemon, 1999; Fornell, 1992; Homburg, Koschat, & Hoyer, 2005). Next, the relationship between customer satisfaction and enhanced firm performance, which stems from the aforementioned favorable customer outcomes, has been reported in a growing number of studies (Bolton, 1998; Grica & Rego, 2005; Mittal & Kamakura, 2001; Rust, Zahorik, & Keiningham, 1994). However, in contrast to the substantial support documented for the first two links in the chain, empirical findings pertaining to the association between customer satisfaction and the response of financial markets are “more mixed” (Anderson & Mansi, 2009, p. 704). In particular, this inconsistent finding is manifested by the lack of a consistent short-run market reaction to customer satisfaction announcements (Fornell et al., 2006; Ittner and Larcker, 1998; Ittner et al., 2009). Given that customer satisfaction has a demonstrated positive impact on customer behavior, and subsequently, firm performance, the lack of a consistent short-run market response to customer satisfaction announcements is puzzling.

1.2. Research conceptualization

To resolve this conundrum, we propose that investors form expectations with respect to customer satisfaction and that they respond only to deviations from these expectations. Such a premise is well-grounded in the literature. Srinivasan, Pauwels, Silva-Risso, and Hanssens (2009) conceptualize and find that investors respond to deviations from expectations for a wide range of marketing investments and actions (e.g., innovation, advertising, price promotions, competitive promotions, consumer liking, and quality). Similarly, Joshi and Hanssens (2009) report that the post-launch performance of studio stock price is strongly influenced by expectations of performance built up prior to the release. Our premise that the stock market responds only to new information is also consistent with research in the finance and accounting literature. For example, in their seminal empirical analysis of market reactions to earnings announcements, Ball and Brown (1968) show that the market responds to the unexpected components of earnings announcements (earnings surprise). Specifically, they demonstrate that the market reacts in the same direction as the difference between actual income and expected earnings. Subsequent researchers (Freeman & Tse, 1992; Kinney, Burgstahler, & Martin, 2002) have also documented the market response to earnings surprises. They find an S-shaped response, with steeply sloped market responses for small absolute surprises and approximately flat market reactions for large absolute surprises.4

At this point, the question arises: How do investors form expectations about the level of customer satisfaction? We posit that expectations of customer satisfaction in the current period are influenced by customer satisfaction in the prior period augmented by marketing investments allocated to improving customer satisfaction in the prior period. Although marketing investments allocated to customer satisfaction in the prior period are not directly observable, we further posit that variables associated with the firm’s environment and its overall marketing investments in the prior period can potentially proxy for these investments in customer satisfaction. For example, firms that face high growth prospects are more likely to invest in customer satisfaction because of the potential to leverage such investments across new opportunities. Similarly, firms that invested in advertising in previous periods will likely exhibit higher levels of customer satisfaction because of the utility-enhancing effects of advertising.

Accordingly, we posit that prior to the release of new customer satisfaction scores, investors have their own private forecasts of the customer satisfaction levels of a particular firm. Moreover, we further predict that the pre-announcement stock price is influenced by these private forecasts. In other words, it is as though some underlying consensus view of all investors is reflected in the pre-announcement stock price.

Next, we hypothesize that the response to customer satisfaction announcements is as follows. Firms that report numbers that are above the underlying consensus forecast earn significantly positive abnormal returns. Conversely, firms that report numbers that are below the underlying consensus forecast earn significantly negative abnormal returns. In effect, the market only responds to the "surprise" or "new" information in the customer satisfaction announcement. Indeed, Ittner et al. (2009) take an important first step in this direction when they look at the market reaction to year-over-year changes in customer satisfaction. However, as argued by Jacobson and Mizik (2009b, p. 842) it is also important to allow market participants to “use other information to adjust their expectations” about customer satisfaction. It is in this way that we build and extend the current literature to better understand the short-run market reaction to customer satisfaction announcements.

1.3. Alternative conceptualizations

Of course, there are alternative conceptualizations for the lack of a consistent, short-term response of the financial markets to announcements of customer satisfaction. It may well be that investors are simply not aware of customer satisfaction announcements. A related view is that investors may be aware of the customer satisfaction metric but ignorant of the many positive impacts of customer satisfaction. For either of these reasons, investors may thus come to disregard customer satisfaction, thereby leading to the observed lack of market reaction.

A second conceptualization is that investors react to customer satisfaction information before the announcement because they are able to track other marketing indicators prior to the announcement. In this regard, Ntgo, Casta, and Ramond (2011) propose that investors do care about customer satisfaction but they can obtain information on customer satisfaction or dissatisfaction well before customer satisfaction scores are publicly announced. For example, as in Tellis and Johnson (2007), investors may respond positively to reviews of product quality; consequently, announcements of improved customer satisfaction stemming from improved quality provide no additional information.

Finally, a third conceptualization is that investors react to customer satisfaction information only after the announcement, when the positive chain of effects resulting from customer satisfaction has, in fact, materialized. Investors may also wait for customer satisfaction information to be filtered through analysts before responding to it. Along these lines, Luo, Homburg, and Wieseke (2010) show that customer satisfaction leads to lower forecast errors from analysts. Because investors generally find it costly to process the information

3 We clarify that in all of the aforementioned studies and in our research endeavor, the conceptual development is around the general construct of customer satisfaction; however, the utilized measure of customer satisfaction is the ACS metric provided by ACSI LLC.

4 Strictly speaking, this body of research does not require individual investors to form expectations; rather, what is required is that investors, in the aggregate, act as though they were responding to deviations from expectations.
surrounding a customer satisfaction announcement, it is thus worthwhile for them to wait for analyst forecasts before responding.

We respond to the first of these alternative conceptualizations by providing three distinct sets of empirical evidence that demonstrate that market participants do indeed care about customer satisfaction. The first set of empirical evidence includes the fact that many prominent industry associations, investment websites, and companies choose to broadcast information about customer satisfaction metrics. The second set of empirical evidence pertains to the increased volume of web searches for customer satisfaction around scheduled customer satisfaction announcements. Finally, the third set of empirical evidence pertains to increased trading volume around customer satisfaction announcements. Taken together, these three sets of empirical evidence refute the notion that market participants do not care about customer satisfaction.

Our response to the other alternative conceptualizations — that investors may respond either before or after the announcement — is somewhat different. Indeed, it is highly likely that some of the response to customer satisfaction information may precede the announcement. It is also highly likely that some of the response to customer satisfaction information will follow the announcement. However, we argue that these effects should only hamper our efforts to find any effect at the time of the announcement. Therefore, any effects that we find can be considered conservative.

The remainder of the paper is organized in the following manner. Given the alternative conceptualization that investors may simply disregard customer satisfaction, we first provide three sets of empirical evidence to demonstrate that investors do indeed follow our empirical metric of customer satisfaction, namely the American Customer Satisfaction Index (ACSI). We then develop our conceptual model of customer satisfaction expectations. Subsequently, we describe our data, variables, and summary statistics. Then, we present our findings pertaining to the two expectations models: a naïve model where expected customer satisfaction is simply last year’s customer satisfaction and a model that includes firm characteristics and marketing investments to proxy for prior resources that were allocated to customer satisfaction. Finally, by examining the market reaction to customer satisfaction announcements, we find that the market does indeed respond to “surprises” in customer satisfaction announcements, with more pronounced effects in our second model of expectations. We conclude with a summary and discussion of contributions and limitations.

2. Are ACSI announcements followed by investors?

An important question to resolve before we proceed is: Do investors follow ACSI announcements? Accordingly, we provide three distinct sets of empirical evidence. First, we note that ACSI announcements are immediately highlighted and re-broadcast by leading industry associations, investment websites, and companies. For example, Progressive Grocer routinely summarizes changes in ACSI scores among food processing companies. In particular, a recent announcement highlights the following improvements: “Sara Lee Corporation showed the biggest improvements in its ACSI score,” “Mars Inc., Nestlé, General Mills Inc. and ConAgra Foods all posted scores of 83,” and “Hershey Foods Corporation edged up a point.”

Similar reports are provided by American Banker in the banking industry, Wards Auto in the automobile industry, Nation’s Restaurant News in the restaurant industry, PC Mag in the computer industry, and Insurance Networking News in the insurance industry. In a similar fashion, ACSI announcements are also highlighted and re-broadcast by several investment websites. For example, Seeking Alpha, a popular investment website recently featured the following headline: “Heinz Ranks Number One in Customer Satisfaction Among All 225 Companies in the American Customer Satisfaction Index.”

Another investment website, www.istockanalyst.com, also provides updates on ACSI scores. Finally, companies themselves are not shy about providing press releases about improvements in customer satisfaction. For example, in a conference call held by Comcast Corporation, Brian Roberts, Chairman and CEO, reiterates his goal of providing a superior customer satisfaction experience while noting a 9% improvement in ACSI scores. Similarly, following an ACSI score announcement in which Yahoo obtained higher scores than Google, a Yahoo spokesperson stated that, “Yahoo is pleased with the results of this year’s ACSI study, which reflect our continued efforts to enhance the consumer experience for more than 500 million users of Yahoo branded properties around the world.” In addition, Sprint Corporation recently highlighted its strong showing in customer satisfaction via press releases and full-page ads in the Wall Street Journal in which the ACSI logo was prominently displayed in tandem with the tag line, “The real reward is making customers happy.”

Second, we examine the intensity of online search for the term “American Customer Satisfaction Index” on Google, the search engine with the highest market share. A number of recent studies (Da, Engelberg, & Gao, 2011; Joseph, Wintoki, & Zhang, 2011) have demonstrated that online search intensity serves as a valid proxy for investor attention. Moreover, this proxy reliably predicts stock returns and trading volume. Using data from Google Insights, where such data are archived, we examine the intensity of searches for the term “American Customer Satisfaction Index” over the period 2004 to 2006 (the time period over which the availability of Google search data overlaps that of our sample). Notably, we find that Google searches for the term “American Customer Satisfaction Index” are seven times higher during the weeks of ACSI score announcements than in other weeks throughout the period. To further illustrate our point, Fig. 1 displays the level of public interest in ACSI scores by plotting the weekly search intensity for the term “American Customer Satisfaction Index” during the year 2006. The displayed search intensity is normalized by a factor that is unknown for privacy and other reasons; consequently, the absolute value displayed is not meaningful. Nevertheless, the figure clearly shows that Google searches for the term spike during the week of, and the weeks surrounding, the announcement of ACSI scores. We interpret this finding as evidence that ACSI announcements are followed by some investors.

Third, we examine changes in trading volume around ACSI announcements. If trading volume increases around ACSI announcements, we can conclude that there is heightened investor attention. To examine this, we calculate the average daily volume for each of our sample firms and for the market during a “non-announcement window”. This window includes a stretch ranging from fifty (t = 50) to fifteen days before the ACSI announcement (t = 15), and another stretch from fifteen (t + 15) to fifty days after the announcement (t + 50), where t represents the date of the ACSI score announcement. We find that the average volume is 2.8% higher in the ten-day period (−5 ≤ t ≤ 5) around ACSI announcements than during the “non-announcement” window, relative to the market.

Based on these three sets of empirical evidence, we thus conclude that customer satisfaction, as represented by the ACSI metric, is closely followed by at least some investors. We next discuss the development of our conceptual model of customer satisfaction expectations.

---

10 The market is defined as the entire Center for Research in Security Prices (CRSP) database. The 2.8% increase in volume is significant at a 10% level using a two-tailed t-test.
3. Model of customer satisfaction expectations

As previously suggested, we posit that the expectations of the level of customer satisfaction at time t, $E[CSt_t]$, are influenced both by the level of customer satisfaction in period $t-1$, $CSt_{t-1}$, and investments in customer satisfaction during period $t-1$, $I_{t-1}$. Formally, we have:

$$E[CSt_t] = \alpha CSt_{t-1} + \beta I_{t-1}.$$  \hspace{1cm} (1)

Eq. (1) suggests that there is some carry-over of the goodwill inherent in customer satisfaction from period to period. This is augmented by an increase in customer satisfaction on account of prior investments in customer satisfaction improvement, $I_{t-1}$.

Now, in forming expectations of customer satisfaction in period $t$, $E[CSt_t]$, investors cannot observe all the myriad investments devoted to customer satisfaction during period $t-1$, $I_{t-1}$. Nevertheless, investors have some beliefs about $I_{t-1}$ based on publicly observable firm characteristics and marketing investments at time $t-1$. We therefore use publicly observable firm characteristics and marketing investments at time $t-1$ to serve as proxies for the resources allocated to customer satisfaction at time $t-1$.

Which variables will investors use to estimate $I_{t-1}$? We consider seven variables: growth opportunities, number of segments served, market concentration, efficiency in converting inputs to outputs, firm size, advertising, and R&D. We next develop the logic associated with our use of these seven variables.\(^\text{11}\)

With respect to growth opportunities, investors believe that firms with high growth prospects are likely to invest more in customer satisfaction at time $t-1$ because they can potentially recoup these investments as these new opportunities are exploited. High levels of customer satisfaction in current businesses can be gainfully leveraged to enter new businesses via lower acquisition costs (Anderson et al., 1994) or used to benefit from a higher receptivity to cross-selling initiatives (Bolton, 1998). Similarly, firms operating in multiple segments will likely invest more in customer satisfaction because of spillover benefits. Chai and Ding (2009) provide empirical evidence for such customer satisfaction spillovers in the mobile phone industry (between handset manufacturers and network operators). Next, because market concentration is a proxy for barriers to entry (Powell, 1996), it follows that firms operating in concentrated industries will have greater incentives to invest in customer satisfaction because they can appropriate more of the value associated with their investments.

We expect efficiency in converting inputs to outputs and firm size to also straightforwardly impact investments in customer satisfaction. Highly efficient firms will see more benefits from such investments while large firms will choose to invest less on account of the heterogeneous customer base that they are likely to serve (Anderson et al., 1994). Finally, we expect investments in advertising and R&D to positively impact customer satisfaction expectations. As demonstrated by Erdem and Sun (2002), advertising can increase the mean and decrease the variance of customer utility, thereby increasing surplus and customer satisfaction. Similarly, we expect investments in R&D to also allow firms to provide greater product differentiation (Chauvin & Hirschey, 1993; Veliyath & Ferris, 1997). This positive differentiation should also enhance utility and increase customer satisfaction.

Clearly, these seven variables do not exhaust all of the influences of firm investments in customer satisfaction improvements. Rather, we make the more modest claim that, as a group, they capture important aspects of the decision to invest in customer satisfaction and serve as good proxies for the information available to investors.

Growth opportunities and the number of segments served reflect the trajectory and depth of market demand for a firm’s product and are therefore intimately related to the future payoffs from customer satisfaction investments. Concentration incorporates the influence of the competitive environment in that it reflects the ability of a firm to appropriate the value that it creates through customer satisfaction investments. Efficiency in converting inputs to outputs and firm size pertain to the firm’s productivity in utilizing its resources. Finally, advertising and R&D expenditures capture firms’ direct investments in customer satisfaction.

We also do not claim that any or all of these variables are exogenous with respect to customer satisfaction. For example, managerial ability could affect both the expected level of customer satisfaction and the market’s assessment of growth opportunities between customer satisfaction announcements. Because this firm-specific effect is not completely observable, failing to address it may introduce an omitted variable bias into the regression estimates. Simultaneity between customer satisfaction and any of the firm characteristics we propose is another issue we acknowledge. For example, in our subsequent empirical analysis, we use market-to-book value as a proxy for growth opportunities. An increase in growth opportunities at time $t-1$ may cause investors to increase their expectations with regard to what customer satisfaction will be at time $t$. However, the reverse could also be true. If investors expect customer satisfaction to be higher next period, they will bid up the market value of the firm, thus increasing its market-to-book ratio.

Finally, we acknowledge the dynamic endogeneity, or reverse causality, of customer satisfaction to our firm characteristics. For example, firms with high growth opportunities are likely to be firms that have had positive shocks to customer satisfaction in the past. In our subsequent analysis, we are careful to account as much as possible for all these sources of endogeneity in our empirical estimation of expected customer satisfaction.

Given the arguments we have presented in this section, we express our model of customer expectations as:

$$E[CSt_t] = \alpha CSt_{t-1} + \beta X_{t-1}.$$  \hspace{1cm} (2)

\(^{11}\) More generally, it is common practice in the finance and accounting literature to predict expected levels of key marketing investments by employing observable firm characteristics. Lou (2011), for example, employs an expected model for advertising investments by utilizing such firm characteristics as market-to-book value, firm size, sales, assets, and capital expenditures. Bushee (1998) employs market-to-book value, firm size, and leverage as control variables in a model that predicts changes in R&D investments. Publicly available information is also often used to predict outcomes related to the firm. For example, Frankel and Lee (1998) and Hughes, Liu, and Wei (2008) utilize firm characteristics to predict analyst forecast errors. The essential idea is that these firm characteristics are reflective of the information set that is available to investors.
where the vector \( X \) encapsulates the seven described variables. Then, announcements of customer satisfaction in period \( t \), \( CS_t \), lead to an investor response that is proportional to \( CS_t - E[CS_t] \).

4. Data, variables, and summary statistics

4.1. Data

We conduct our analysis within the sample of firms for which there is satisfaction (ACSI) data between 1995 and 2006. As described in Anderson et al. (2004), the ACSI methodology provides a "uniform, independent, customer-based, firm-level satisfaction measure for nearly 200 companies in 40 industries and in seven sectors of the US economy" (p. 176).

Because our explanatory variables (described below) are from the COMPUSTAT database, we merge the ACSI and COMPUSTAT data. This leaves us with a working sample consisting of 116 firms and 1109 firm-years over the 12-year period from 1995 to 2006.

4.2. Variables

We construct proxies for our independent variables as follows. With respect to an empirical proxy for future growth, we note that it is quite common to employ the market-to-book ratio as a measure of the growth opportunities associated with the firm in the finance literature (e.g., Denis, 1994; p. 162). We compute market-to-book value as the market value of assets divided by the book value of assets, where the market value of assets is computed as the book value of assets (COMPUSTAT item: "ar") plus the market value of common stock (the product of COMPUSTAT items: common shares outstanding, "csho" and fiscal year end stock price, "prcc_f") less the sum of the book value of common stock (COMPUSTAT item: "ceq") and balance sheet deferred taxes (COMPUSTAT item: "txdbf"). Next, the number of business segments is obtained by counting the number of business segments from the COMPUSTAT Segments database. With respect to concentration, we use the Herfindahl index (sum of square-market shares) to obtain a continuous measure of industry concentration. Market share, in turn, is computed via sales revenue (COMPUSTAT item: "sale") reported at the 2-digit SIC code level for all COMPUSTAT companies with the same 2-digit SIC code.12

We use return on assets to measure efficiency, computed as operating income before depreciation (COMPUSTAT item: "oibdp") divided by the book value of assets (COMPUSTAT item: "ar"). The proxy for firm size is the log of firm assets, represented by the book value of assets. Finally, advertising is the ratio of advertising expenses (COMPUSTAT item: "xad") to sales (COMPUSTAT item: "sale") and R&D is the ratio of R&D expenses (COMPUSTAT item: "xrd") to sales. Per convention, we set the R&D expense to zero if it is missing or not reported because the SEC has long required all publicly traded firms to report any "material" R&D expenditures (Bound, Cummins, Griliches, Hall, & Jaffe, 1984). Similarly, we set the advertising expense to zero if it is missing or not reported. Firms generally do not have much latitude with respect to disclosing R&D and advertising expenditures; the Generally Accepted Accounting Principles (GAAP) require all firms with "material" R&D or advertising expenditures to recognize and disclose these items in their financial statements. Financial Accounting Standards Board (FASB) Statement No. 2, "Accounting for Research and Development Costs" (1974), sets the standards for R&D accounting; FASB Statement of Position (SOP) 93-7 (1993), "Reporting on Advertising Costs", sets the standards for advertising accounting. This leads us to believe that missing values for R&D and advertising expenses indeed represent zero or close to zero expenditures.

Of course, it is possible that some firms may have advertising expenses that they do not recognize; the aforementioned standards do allow for some exceptions and the very definition of "material" may, on the margin, be subject to auditor judgment. This may make our advertising measure noisy, but a number of studies (e.g., Bound et al., 1984; Chauvin & Hirschey, 1993; Hirschey, Skiba, & Babajide Wintoki, 2012) test the assumption of setting the value of missing R&D and advertising expenses to zero in large samples and conclude that it generally has little to no effect on empirical analyses involving R&D and advertising expenditures.

4.3. Summary statistics

Table 1 displays the summary statistics for the customer satisfaction and firm-specific variables that we use in our analysis. In our sample of 1109 observations, the mean level of customer satisfaction (100 point scale) is 75.3 with a median of 75 and a standard deviation of 6.47. The minimum value of this variable is 49 while the maximum value is 90.

Table 1 also shows that the mean (median) market-to-book ratio is 1.91 (1.41) and the mean or median firm in our sample has approximately 3 business segments. The mean (median) level of concentration is 0.06 (0.04). Return on assets has a mean (median) of 0.14 (0.13) and the mean and median firm sizes, as measured by the log of assets, are 9.67 and 9.63, respectively. Finally, Table 1 shows that the mean (median) advertising-to-sales ratio is 3% (1%) and the mean (median) of R&D-to-sales is 1% (0).

Table 2 displays the correlation matrix between customer satisfaction and other firm-specific variables. We find that customer satisfaction is positively correlated with the market-to-book ratio, number of business segments, return on assets, advertising-to-sales ratio, and R&D-to-sales ratio. It is negatively correlated with firm size.

5. Empirical model and findings

We organize our findings into two main parts. First, we estimate our proposed models of customer satisfaction expectations. In estimating our models, we employ both OLS estimation as well as the dynamic panel generalized method of moments estimation. As we will discuss later, the latter method accounts for potential sources of endogeneity. We then examine the market's reaction to customer satisfaction surprises, followed by additional robustness tests.

5.1. Empirical model and estimation of expected customer satisfaction

Our basic model for examining the determinants of expected customer satisfaction is:

\[
CS_{it} = \alpha_0 + \alpha_1 CS_{it-1} + \beta_1 X_{it-1} + \eta_t + d_t + e_{it}
\]  

(3)

where \( CS_{it} \) is the customer satisfaction of firm \( i \) in year \( t \), \( X_{it-1} \) is a vector of observable firm characteristics and investments (market-to-book value, the number of business segments, concentration, return on assets, firm size, advertising, and R&D expenditures) that proxy for the firm's prior investments in customer satisfaction, \( \eta_t \) is an unobservable firm-fixed effect, \( d_t \) is a year dummy, and \( e_{it} \) is an error term. The firm-fixed effect is especially important because it controls for any other unobservable firm characteristics that may be important in determining the expected level of customer satisfaction. The observable firm characteristics are all lagged — they are measured at the end of the fiscal year prior to the customer satisfaction announcement.

Although we start our empirical analysis with a simple OLS estimation of Eq. (3), there are at least three sources of endogeneity that could potentially bias our OLS estimates. Hence, we apply the dynamic panel data estimation methodology to obtain System GMM.

12 The Herfindahl Index is widely utilized as a proxy for industry concentration (Huo and Robinson, 2006).
The coefficient of lagged customer satisfaction is 0.91. This high level of persistence raises the possibility of a unit root in customer satisfaction. However, we are able to reject the null hypothesis of a unit root through a Fisher test for panel unit roots, using an augmented Dickey–Fuller test.

estimates. Technical details of the various sources of endogeneity and dynamic panel data estimation method are presented in Appendix A.

Both OLS and System GMM results are shown in Table 3. Because we are using panel data for our empirical analysis, it is possible that the residuals for a given firm can be correlated across years (serial correlation), or that the residuals of a given year may be correlated across different firms (cross-sectional dependence); both of these could bias the standard errors downwards and inflate the t-statistics (Petersen, 2009). Thus, to avoid biased inferences, we include year dummies in all specifications and report t-statistics that are clustered by firm and industry in all our regression results.

The first column in Table 3 shows the results of regressing this year’s customer satisfaction on last year’s customer satisfaction. The estimate clearly shows that lagged customer satisfaction is persistent; the coefficient of lagged customer satisfaction is 0.91. This finding has previously been documented in the literature (see, for example, Tuli & Bharadwaj, 2009) and is likely due to a combination of two factors: (i) the persistence of customer satisfaction arising, as we propose, from the fact that there is some carry-over of the goodwill inherent to customer satisfaction from period-to-period, and (ii) the presence of an unobservable fixed effect that is correlated both with customer satisfaction and the firm-specific variables (Xit−1) included in Eq. (3).

Next, in column (2), we report the results of estimating a model that includes both firm characteristics and the lagged value of customer satisfaction using OLS. There are two noteworthy points. First, the R² improves from 81% to approximately 84%, and this increment is statistically significant. Second, the Vuong (1989) Z-statistic that compares the model reported in column (2) to the model reported in column (1) is 5.42 (p < 0.01). In addition, the F-statistic for firm characteristics and marketing investments is a statistically significant 2.12 (p = 0.04). These findings suggest that, in addition to past customer satisfaction, the firm characteristics and marketing investments that we have chosen to proxy for previous investments in customer satisfaction provide additional explanatory power for the cross-sectional variation in realized customer satisfaction.

While the OLS estimates provide some initial support for our theoretical conjecture that expected customer satisfaction will be a function of past customer satisfaction and proxies for the firm’s informational environment, these estimates do not effectively address unobservable heterogeneity, simultaneity, and reverse causality. For this we turn to the GMM estimates in column (3) of Table 3.
Table 3
Regression results. The dependent variable in all specifications is the customer satisfaction (ACSI score). Market-to-book value is the market value of assets divided by the book value of assets where the market value of assets is computed as the book value of assets plus the market value of common stock less the sum of the book value of common stock and balance sheet deferred taxes. Log segments is the log of the number of business segments, which is obtained by counting the number of business segments from the COMPUSTAT Segments database. Concentration is the Herfindahl index (sum of square-market shares) where market share is computed via sales revenue reported at the 2-digit SIC code level for all companies with the same 2-digit SIC code. Return on assets is computed as the operating income before depreciation divided by the book value of assets. Log assets is the log of firm assets, represented by the book value of assets. Advertising is the ratio of advertising expenses to sales, R&D is the ratio of R&D expense to sales, and industry, are in parentheses. P-Values, where appropriate, are in brackets.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market-to-book</td>
<td>0.11***</td>
<td>0.23**</td>
<td>System GMM</td>
</tr>
<tr>
<td>Log (segments)</td>
<td>(2.77)</td>
<td>(2.60)</td>
<td></td>
</tr>
<tr>
<td>Concentration</td>
<td>0.08</td>
<td>0.34*</td>
<td></td>
</tr>
<tr>
<td>ROA</td>
<td>0.16</td>
<td>8.53***</td>
<td></td>
</tr>
<tr>
<td>Log (assets)</td>
<td>(0.19)</td>
<td>(2.65)</td>
<td></td>
</tr>
<tr>
<td>Advertising</td>
<td>4.51***</td>
<td>6.19*</td>
<td></td>
</tr>
<tr>
<td>R&amp;D</td>
<td>5.30</td>
<td>18.74</td>
<td></td>
</tr>
<tr>
<td>Lagged customer satisfaction</td>
<td>0.91***</td>
<td>0.89**</td>
<td>(1.32)</td>
</tr>
<tr>
<td>Constant</td>
<td>6.35</td>
<td>7.83**</td>
<td>(1.12)</td>
</tr>
<tr>
<td>Time dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.8125</td>
<td>0.8355</td>
<td></td>
</tr>
<tr>
<td>Vuong Z-stat of (2) vs. (1) p-value</td>
<td>5.42***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-stat of firm characteristics p-value</td>
<td>2.12**</td>
<td>10.05***</td>
<td></td>
</tr>
<tr>
<td>AR (2) test p-value</td>
<td>[0.04]</td>
<td>[0.01]</td>
<td>[0.01]</td>
</tr>
<tr>
<td>Hansen test of over-identification p-value</td>
<td>[0.99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of firms (firm-years)</td>
<td>115 (949)</td>
<td>115 (949)</td>
<td>115 (949)</td>
</tr>
</tbody>
</table>

*** Represent significance at the 1% level, using two-tailed tests.  
** Represent significance at the 5% level, using two-tailed tests.  
* Represent significance at the 10% level, using two-tailed tests.

The System GMM results in column (3) of Table 3 provide evidence that past customer satisfaction and our proxies for the firm’s prior investments in customer satisfaction explain the cross-sectional variation in customer satisfaction. Individually, we find that customer satisfaction is positively associated with market-to-book value, number of segments, return on assets, and advertising expenditures. Even more importantly, the firm characteristics and marketing investments are jointly significant at the 1% level (F-statistic = 3.05).

Table 3 also shows the results of two post-estimation tests of the validity of the dynamic GMM specification as well as the validity of the instrument set. The first test is a test of serial correlation. Table 3 shows the results of an AR(2) test of the null hypothesis of no second order serial correlation. The results of this test confirm that this is the case: the AR(2) test yields a p-value of 0.14. The second test is a Hansen test of over-identification. The Hansen test yields a J-statistic that is distributed χ² under the null hypothesis of the validity of our instruments. The results in Table 3 reveal a J-statistic with a p-value of 0.99.

Overall, the results in Table 3 suggest that our proposed empirical model provides a fair approximation of investor expectations of customer satisfaction. Next, we turn to examining the cross-sectional relation between the abnormal stock returns around customer satisfaction announcements and deviations from “expected” customer satisfaction. Of course, these deviations reflect “truly” unexpected changes in customer satisfaction and “errors” due to model misspecification (including missing variables, incorrect functional forms, and exclusion of trends); however, the relatively high goodness-of-fit measures suggest that the latter is not a serious concern.

5.2. Market reaction to surprises in customer satisfaction announcements

In Table 4, we report the cumulative abnormal return by various surprise quintiles, wherein firms are arranged into quintiles based on the difference between actual customer satisfaction and expected customer satisfaction. Quintile 1 consists of firms having the most negative difference between actual customer satisfaction and the predicted (expected) level of customer satisfaction, while Quintile 5 contains those with the most positive difference. The expected customer satisfaction is obtained via two models: naïve (only utilizes last year’s level of customer satisfaction) and GMM (utilizes both a lagged value of customer satisfaction and a vector of lagged firm-characteristics and marketing investments) as reported in column (3) of Table 3. Thus, for the naïve model, the “surprise” Δit for firm i at time t is given simply by Δit = CSit − CSit−1, while for the GMM model, the surprise is given by Δit = CSit − α0 − α1CSit−1 − 1β0Yit−1 − 1ηit − δi. In Table 4, we report six-day (0, +5) returns across two risk adjustments: (i) market-adjusted, and (ii) market-, size-, book-to-market-, and momentum-adjusted. In each case, the cumulative abnormal return (CAR) is obtained by summing the abnormal return (AR) of each firm over the event window and then averaging across the firms in the quartile. Thus, for each firm i,

\[ \text{CAR}_i = \sum_{t=-5}^{t=0} \text{AR}_{it}. \]

The precise calculations of the abnormal returns for the market-adjusted model and for the market-, size-, book-to-market- and momentum-adjusted model are detailed in Appendix B.

Focusing on the market-adjusted returns in Table 4, we find that the naïve model (model 1) demonstrates a significant abnormal reaction (1.05%, t = 2.99) only for large positive surprises in customer satisfaction. This is consistent with results reported in Ittner et al. (2009). In contrast, the GMM model demonstrates a significant abnormal reaction to both large positive surprises (1.21%, t = 2.24) and large negative surprises (−0.74%, t = −2.06). For both models, the difference between Quintile 5 and Quintile 1 is statistically significant; however, the difference between Quintile 5 and Quintile 1 is bigger in the GMM model (1.95%, t = 3.61) than in the naïve model (1.64%, t = 3.32). These results demonstrate that the market does indeed respond in the short-term to surprises in customer satisfaction, with substantially more pronounced effects in the GMM expectations models.14 The findings are very similar when the abnormal returns are market-, size-, book-to-market- and momentum-adjusted. Finally, Quintiles 2, 3, and 4 are characterized by less extreme levels of surprise and we find little reaction to those announcements.

To obtain an idea of the economic significance of the cross-sectional variation of abnormal returns, we note that the average customer satisfaction surprise in Quintile 1 is −3.8, while that in Quintile 5 is 3.4; a difference of 7.2 customer satisfaction points. Again, if we consider the six-day (0, +5) cumulative abnormal reaction (using the GMM expectations model, and market, size, book-to-market and momentum adjustment for the returns in Table 4), the difference in

14 All t-statistics in this table are based on standard errors that account for industry and event-date clustering.
returns between Quintiles 5 and 1 is 2.16%. The average firm in our sample has a market capitalization of $2.96 billion. Thus, 2.16% represents approximately $64 million in market capitalization for the average firm, suggesting that each “unexpected” customer satisfaction point is worth $9 million to the typical firm.

To further put this finding in context, assume that a firm is currently in the 25th percentile of customer satisfaction (with a score of 71) and is considering improving its score to reach the 75th percentile of customer satisfaction (a score of 80). Our findings reveal that this 9-point improvement in customer satisfaction will increase the market value of the firm by approximately $80 million. These estimates provide some broad cost–benefit guidelines for investments in customer satisfaction. Specifically, initiatives that increase customer satisfaction by 9 points or prevent an anticipated 9-point decline in customer satisfaction are worthwhile as long as they do not exceed $80 million.

5.3. Robustness tests

We carry out a number of robustness tests of the analysis in Table 4. In Table 4, we have used a six-day (0, +5) window around the customer satisfaction announcement. However, it is possible that there might be significant leakage of customer satisfaction information prior to the announcement date we have identified, and this might be driving our results. In Table 5, we replicate the analysis in Table 4 for a longer window — eleven days (−5, +5) around the announcement. The results in Table 5 show that, even with the longer window, there is a positive relation between the announcement return and the magnitude of the surprise. Again, we see that this positive relation is more pronounced in the full GMM model than in the naïve model. The results also show that the returns (and the pattern of returns) in the longer window are of a similar magnitude to those in the (0, +5) window.

In Fig. 2, we summarize the CARs across the entire eleven-day window, from 5 days before to 5 days after (−5, +5) the announcement. The figure shows the difference between Quintile 5 and Quintile 1 with the abnormal returns being market-, size-, book-to-market- and momentum-adjusted. The figure shows that there is very little pre-announcement leakage; the difference in abnormal returns between Quintile 5 and Quintile 1 is insignificant in the 5 days (−5, −1) before the identified announcement date. This is in clear contrast to the significant difference between Quintile 5 and Quintile 1 in the 5 days (0, +5) following the announcement of customer satisfaction scores.

In an untabulated analysis, we examine the robustness of our surprise measure by excluding concentration, log of assets, and R&D expense from our calculation of expected customer satisfaction. These are variables that are not individually significant in the GMM regression in column (3) of Table 3. The abnormal returns we obtain are similar to those we report in Table 4. For example, the (market-adjusted) difference in returns between Quintile 5 and Quintile 1 is 2.05% (t = 3.60), while for the market-, size-, book-to-market- and momentum-adjusted returns, the difference is 1.88% (t = 3.32).

Finally, in an additional untabulated analysis, we estimate Eq. (3) using a Box–Cox transformation and calculate our surprise measure using the predicted values. This allows estimation without a priori specification of the functional form. We then examine the cumulative abnormal return by various surprise quintiles. We find results that are qualitatively and quantitatively similar to those obtained by employing the naïve and GMM models. The abnormal return for the lowest quintile (Quintile 1) is negative but insignificant, and the abnormal return for the highest quintile (Quintile 5) is positive and significant. The (market-adjusted) difference in returns between Quintile 5 and Quintile 1 is 1.51% (t = 3.06), while for the market-, size-, book-to-market- and momentum-adjusted returns, the difference is 1.82% (t = 3.37). This suggests that our GMM results are not driven by the assumption of a linear model specification.

6. Summary, contributions and limitations

6.1. Summary and contributions

Our research is motivated by the lack of evidence pertaining to short-run investor reactions to customer satisfaction announcements. Clearly, this gap diminishes the received conceptualization of customer satisfaction. Accordingly, we set out to assess short-run investor...
Table 5
Cumulative abnormal returns (CAR) and unexpected changes in customer satisfaction around a longer event window. The table reports the cumulative abnormal reaction (CAR) to the announcement of customer satisfaction scores in a (-5, +5) window around the announcement. Day 0 is the date of the announcement of the customer satisfaction scores. Firms are arranged into quintiles based on the difference between the actual and expected customer satisfaction scores, with Quintile 1 representing the most negative difference and Quintile 5 representing the most positive difference. For Model 1, the expected customer satisfaction is \( \text{CS}_{t-1} \), while for Model 2, the expected customer satisfaction is \( \alpha - \beta_{1} \text{CS}_{t-1} - \beta_{2} \text{X}_{t-1} \), where \( \text{X}_{t-1} \) is a vector of variables that includes market-to-book value, number of market segments, concentration, industry, return on assets, firm size, advertising, and R&D. 

<table>
<thead>
<tr>
<th>Quintile (Q)</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>All firms</td>
<td>0.22%</td>
<td>0.35%</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(1.54)</td>
</tr>
<tr>
<td>1</td>
<td>-0.05%</td>
<td>-0.12%</td>
</tr>
<tr>
<td></td>
<td>(-0.09)</td>
<td>(-0.22)</td>
</tr>
<tr>
<td>2</td>
<td>-0.83%</td>
<td>-0.32%</td>
</tr>
<tr>
<td></td>
<td>(-1.23)</td>
<td>(-0.71)</td>
</tr>
<tr>
<td>3</td>
<td>-0.60%</td>
<td>-0.06%</td>
</tr>
<tr>
<td></td>
<td>(-1.37)</td>
<td>(-0.12)</td>
</tr>
<tr>
<td>4</td>
<td>0.39%</td>
<td>0.61%</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>5</td>
<td>1.73%***</td>
<td>1.98%***</td>
</tr>
<tr>
<td></td>
<td>(3.30)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>Q5 minus Q1</td>
<td>1.79%**</td>
<td>2.10%**</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(2.50)</td>
</tr>
</tbody>
</table>

*** Represent significance at the 1% level, using two-tailed tests.

** Represent significance at the 5% level, using two-tailed tests.

* Represent significance at the 10% level, using two-tailed tests.

reactions to customer satisfaction announcements. A key conceptual contribution of our research efforts is the incorporation of the well-developed notion of investor expectations into the customer satisfaction context. In our work, we analyze two distinct “expectations” models: a naïve model where expected customer satisfaction is simply last year’s customer satisfaction, and a model that includes firm characteristics and marketing investments. For both of these “expectations” models, we utilize prediction errors to obtain a measure of customer satisfaction “surprise.” We then examine the market’s response to our customer satisfaction surprise measure. In doing so, we find that the market does indeed respond in the short-term to surprises in customer satisfaction, with more pronounced effects for our GMM model. This empirical finding is a second contribution of our research efforts.

Our empirical findings differ from those reported in Fornell et al. (2006) and Ittner and Larcker (1998) but complement those documented in Ittner et al. (2009). We extend the naïve expectations model analyzed by Ittner et al. (2009), motivated by the suggestions in Jacobson and Mizik (2009b). Specifically, we include firm characteristics and marketing investments and utilize GMM to explicitly control for unobserved heterogeneity, simultaneity, and dynamic endogeneity. Controlling for these various sources of endogeneity allows us to estimate the true model with greater confidence. Strikingly, we find that these additional variables merit inclusion from a statistical perspective. As such, they give credence to our enhanced expectations model.

6.2. Limitations

Of course, our work is not without limitations. First and foremost, developments in the theory of investor expectations may suggest other variables that warrant inclusion in a model of customer satisfaction expectations. Thus, we can make no claim that we have developed the “best” model of customer satisfaction expectations; rather, we view our development of a model of customer expectations as an important first step in understanding this phenomenon. Nevertheless, the statistical significance of the variables in our expectations model, along with its ability to demonstrate the anticipated short-term effects, reveals the inherent logic and contribution of our expectations model.

Second, we have developed and analyzed a model of customer satisfaction expectations across firms in different industries. It may be possible to build models that are specific to firms in particular sectors, thereby allowing for an expectations model with a somewhat more comprehensive set of predictor variables, such as investments in training and service infrastructure improvements. Such an analysis may yield fairly precise estimates of customer satisfaction expectations. This is indeed the suggestion offered by Jacobson and Mizik (2009b). We hope that better data availability and future research efforts will overcome these limitations.
Appendix A. Sources of endogeneity and dynamic panel data estimation methodology

There are three sources of endogeneity.

Unobservable heterogeneity

The OLS estimation assumes that there are no unobservable firm characteristics that may explain customer satisfaction and that are also correlated with any of our explanatory variables, i.e., it assumes that \( E(\epsilon_t | X_{t-1}) = 0 \). However, as we have argued in developing our conceptual model of expected customer satisfaction, this is unlikely to be the case.

Simultaneity

The OLS estimation assumes that there is no correlation between any of the explanatory or control variables and the error term in Eq. (1), i.e., \( E(\epsilon_t | X_{t-1}) = 0 \). Again, this may not necessarily be the case.

Dynamic endogeneity, or reverse causality

The OLS estimation assumes that the current levels of the explanatory or control variables are independent of previous shocks to customer satisfaction. This is a particularly strong assumption and one that we have argued is unlikely to hold because it implies that all our explanatory and control variables are random draws through time and do not depend on the firm’s history.

To effectively address these sources of endogeneity, we apply the dynamic panel data estimation methodology originally developed by Arellano and Bond (1991) and further developed in a series of papers by Arellano and Bover (1995) and Blundell and Bond (1998).

The basic dynamic panel consists of two key steps. First, we take the first-differences of Eq. (3) to eliminate the fixed effects:

\[
\Delta \text{CS}_t = \alpha_0 + \alpha_1 \Delta \text{CS}_{t-1} + \beta_1 \Delta X_{t-1} + \Delta \epsilon_t
\]  

(A1)

Next, we estimate Eq. (A1) via the general method of moments (GMM), using the explanatory variables that are lagged by two periods \((t-2)\) or more, as instruments for the explanatory variables (which, as we show in Eq. (A1), are measured at \(t-1\)). The first step removes the omitted variable bias that may arise due to (fixed) unobserved heterogeneity. The last step (coupled with the fact that we also include \( \Delta \text{CS} \) at time \( t-1 \) in our basic specification) ameliorates any biases due to simultaneity or dynamic endogeneity.

The basic idea here is that any effect of historical \( \text{CS} \) at time \( t-2 \) or earlier only affects current \( \text{CS} \) (\( \Delta \text{CS} \)) through its effect on \( \Delta \text{CS}_{t-1} \) and \( X_{t-1} \).

Arellano and Bover (1995) and Blundell and Bond (1998) argue that we can improve the GMM estimator by also including the equation in levels in the estimation procedure. We can then use the first-differenced variables as instruments for the equations in levels in a “stacked” system of equations that includes the equations in both levels and differences. This produces a “system” GMM estimator that involves estimating the following system:

\[
\begin{align*}
\Delta \text{CS}_t & = \alpha_0 + \alpha_1 (\Delta \text{CS}_{t-1}) + \beta_1 (\Delta X_{t-1}) + \epsilon_{it} \\
\text{CS}_t & = \alpha_0 + \alpha_1 (\text{CS}_{t-1}) + \beta_1 (X_{t-1}) + \epsilon_{it} 
\end{align*}
\]  

(A2)

Appendix B. Calculation of abnormal returns for the market-adjusted model and for the market-, size-, book-to-market- and momentum-adjusted model

For the market-adjusted model, the abnormal return for each firm is calculated as

\[
\text{AR}_{it} = R_{it} - \alpha - \beta_{im} R_{mt} \tag{B1}
\]

where \( R_{it} \) is the return of firm \( i \) on day \( t \), \( R_{mt} \) is the equal-weighted market return, and \( \{\alpha, \beta_{im}\} \) are the parameters computed by regressing the firm’s returns on market returns from 160 to 10 days before the announcement of customer satisfaction (i.e., a \([-160, -10]\) window relative to the announcement date).

The calculation of abnormal returns for the market-, size-, book-to-market- and momentum-adjusted model is based on the expected returns predicted by a four-factor model. The model consists of three factors from Fama and French (1993): the excess return on the market \((R_m - R_f)\), the return difference between a portfolio of “small” and “big” stocks \((\text{SMB})\) and the return difference between a portfolio of “high” and “low” book-to-market stocks \((\text{HML})\), augmented with a momentum factor from Carhart (1997), which is the return difference between a portfolio of stocks with high returns in the past year and a portfolio of stocks with low returns in the past year \((\text{UMD})\).

Again, the parameters of the expected returns model are computed for an estimation period stretching from 160 to 10 days before the announcement of customer satisfaction (i.e., a \([-160, -10]\) window relative to the announcement date). So for each observation in the sample, the four-factor model parameters are estimated from the regression:

\[
R_{it} - R_f = \alpha_i + \beta_{im} (R_{mt} - R_f) + \beta_{i2} \text{SMB}_t + \beta_{i3} \text{HML}_t + \beta_{i4} \text{UMD}_t + \epsilon_{it} \tag{B2}
\]

We then apply the four factor model parameters obtained from Eq. (B2) to calculate the abnormal returns for each of our event windows:

\[
\text{AR}_{it} = R_{it} - \{\alpha_i + \beta_{im} (R_{mt} - R_f) + \beta_{i2} \text{SMB}_t + \beta_{i3} \text{HML}_t + \beta_{i4} \text{UMD}_t\} \tag{B3}
\]

References


Da, Zhi, Engelberg, Joseph, & Gao, Pengjie (2011). In search of attention.


Making sense of numbers: Effects of alphanumeric brands on consumer inference

Dengfeng Yan a,⁎,1, Rod Duclos b,1

a Department of Marketing, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, United States
b School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

1 Both authors contributed equally to this article.

Abstract

This research examines when and how the presence of seemingly innocuous, non-diagnostic numbers in brand names (e.g., 7-UP) impacts consumers’ judgments. Building on anchoring theory, our central proposition is that numbers contained in alphanumeric brand names can act as implicit anchors that subsequently bias (either upward or downward) consumers’ assessment of a product’s price, weight, volume, etc. We qualify this proposition, however, by showing that such anchoring effects occur primarily when (a) the numeric component of a name appears relevant to the judgment at hand and (b) consumers evaluate product attributes superficially (rather than systematically).

Keywords:
- Alphanumeric brand names
- Anchoring
- Heuristic processing

1. Introduction

Brand and product names play an important role in marketing (Aaker, 1991). Accordingly, brand naming research grew considerably in the last two decades (Klink, 2000; Lowrey & Shrum, 2007; Pan & Schmitt, 1996). Recent findings indicate that brand names can help identify manufacturers and promote products (Friedman, 1985), assist advertising recall (Keller, Hecker, & Houston, 1998), shape product evaluation (Yorkston & Menon, 2004), affect product demand (Sullivan, 1998), and signal quality (Brucks, Zeithaml, & Naylor, 2000). Historically, however, much of the research in the area was guided by a “linguistic” approach wherein the properties of brand names examined were sounds and meanings. Although this paradigm provides an appealing theoretical basis for much brand naming research, its ability to serve as a universal theoretical framework is limited because it is primarily applicable to alphabetic brand names, which neglects the fact that a fair amount of brand names now consist of letters primarily applicable to alphabetic brand names, which neglects the fact that a fair amount of brand names now consist of letters and numbers (e.g., Airbus A330, 7-UP, V8 Juice, And1 Basketball, 7-Eleven, A1 Steak Sauce). Addressing this gap in the literature, Gunasti and Ross (2010) recently showed that consumers hold implicit beliefs about the numbers contained in product names (i.e., greater numbers signal better configuration or quality). As a result, a computer named X-200 is often judged more favorably than one named X-100. Of note, Gunasti and Ross’ (2010) work examines contexts in which consumers are exposed to various options simultaneously. In their study, participants were able to compare products side by side before making a decision. Colloquially, this form of decision-making based on joint evaluations is akin to a “within-subjects” situation wherein consumers can conveniently experience several options at once.

The present work extends the aforementioned line of research by examining decision contexts in which consumers consider products in isolation. We ask, how do alphanumeric brand names impact perceptions, attitudes, and judgments when consumers evaluate a product on its own (i.e., without any point of comparison)? After all, it is quite common for consumers to experience only one product/service at a time (e.g., in advertising, when receiving a gift). As such, what inferences do consumers draw when offered to purchase a computer called X-200? Referring to the earlier colloquialism, our work examines decision-making based on separate product evaluations, which is more akin to a “between-subjects” situation wherein consumers cannot make simple comparisons between products in the marketplace.

With this critical distinction in mind, the central premise of our work is that numbers contained in brand names can be utilized by consumers as anchors to assess product attributes (e.g., price, weight, volume). Furthermore, we propose that alphanumeric brand names influence not only product evaluation and preference-formation but also memory and inference-making. For instance, consumers may perceive that the Airbus A330 has roughly 330 seats and that a can of 7-UP sells for approximately HK$7 (i.e., US$ 0.9) at convenience stores in a city like Hong Kong. Moreover, we argue that anchoring occurs primarily when the number is relevant to (i.e., consistent with) a given product’s attributes and when consumers use a less effortful (i.e., superficial) processing mode, rather than a rigorous (i.e., systematic) one.

In subsequent sections, we first review the literature in this field, highlighting the differences between our research and some recent work on alphanumeric brand names. We then present the results of
five experiments intended to examine our key hypothesis and its boundary conditions. We conclude by discussing the theoretical and managerial implications of our findings.

2. Conceptualization

2.1. Literature review

As mentioned earlier, most research on brand naming focuses on alphabetic brand names by examining their semantic meaning and phonetic characteristics. Past research suggests, for instance, that a brand name should be meaningful and suggestive (Kashmiri & Mahajan, 2010; Lee & Ang, 2003). Robertson recommends brand names that are verbally related to their product class, eliciting mental images, and make use of morphemes (Robertson, 1989). This suggestiveness principle was later qualified by Keller et al. (1998) who argue that a suggestive brand name can also inhibit recall of subsequently advertised product benefits unrelated to the meaning of the brand name. Relatively, Klink (2000) argues that the sound of a brand name (i.e., its phonetic features) can communicate information about a product, such as its size, speed, and weight. Similarly, Yorkston and Menon (2004) show that consumers use the information they gather from phonemes in brand names to infer product attributes and evaluate brands. Lastly, Pan and Schmitt (1996) conducted a comparative study to show that native English speakers' attitudes toward a product are more influenced by sound matching (e.g., a female voice for women's products and a male voice for men's products) than native Chinese speakers.

As mentioned earlier, this linguistic line of research is primarily relevant to purely alphabetic brand names. However, a number of brands in the marketplace now include a combination of letters and numbers (e.g., 3M, 5th Avenue, 67 Jeans, V8 Juice, And1, Airbus A380). Surprisingly, few papers have attempted to rigorously tackle the question of how consumers process alphanumeric brand names. Pavia and Costa (1993), for instance, find that alphanumeric brand names are more suitable for technically complex, manufactured items such as electronics, computers, stereo components, and cameras. In addition, they find that alphanumeric brand names are appropriate for unemotional, formulated products such as vitamin-rich cereals. Ang (1997) investigated how consumers respond to alphanumeric brand names from a socio-cultural perspective. The central idea here is that the same number can have different associations in different cultures. For example, “7”, a lucky number in Western societies, is undesirable in Chinese culture because it is associated with the cult of life after death and death festivities. Recently, Gunasti and Ross (2010) contributed significantly to this slim line of research by showing that consumers implicitly perceive that higher numbers signal better configuration or quality (e.g., a PC called X-200 is more favorable than one called X-100). Qualifying this claim, Gunasti and Ross (2010) also showed that this “higher is better” heuristic is used more (less) by people with a low (high) need for cognition.

2.2. Alphanumeric brand names as implicit anchors

To augment this line of research, we examined the effects of alphanumeric brand names in a new context and from a new perspective. Gunasti and Ross (2010) investigate consumption scenarios in which participants can conveniently evaluate products (whose brand names have been manipulated) side by side before finally choosing one. In contrast, we investigate situations in which consumers evaluate only one product at a time (e.g., viewing marketing communications, receiving a gift). For example, we ask, what inferences would a consumer make from an ad for an MP3 player called M-200? In the absence of a reference point (e.g., M-100), it would be difficult for consumers to rely on Gunasti and Ross’ (2010) “higher is better” heuristic because the very perception of “higher” (or “lower”) requires the presence of a comparison. For situations in which consumers evaluate a product/service in isolation, we propose that the number contained in a brand name may function as an implicit anchor that subsequently biases consumer judgments. The anchoring and adjustment literature (Tversky & Kahneman, 1974) provides the theoretical basis for our argument.

According to Tversky and Kahneman (1974), when a complex task is divided into simpler steps, partial computation often serves as an anchor. For example, Tversky and Kahneman (1974) asked participants to estimate the product of 8 numbers from 1 to 8. These numbers were presented in either ascending or descending order. Students shown “1×2×3×4×5×6×7×8” made a median estimate of 512, whereas those shown “8×7×6×5×4×3×2×1” made a median estimate of 2250. The interpretation of this result is that respondents tried to multiply the first few factors of the product before adjusting either upward or downward. Adjustments were generally insufficient relative to the true value of 40,320. In the first set of guesses, however, adjustments were much more insufficient because they started from a lower anchor.

In a separate study, Tversky and Kahneman (1974) asked students to estimate what percentage of African countries were members of the United Nations. Before they made their estimate, participants were given a random anchor generated by a spinning wheel that contained the numbers 0–100. The wheel was rigged so that half of the participants received 10 as their anchor and the other half 65. Participants were then asked whether the percentage of African countries was higher or lower than their given anchor. Those whose anchor equaled 10 provided an average estimate of 25%, whereas those whose anchor equaled 65 provided an average estimate of 45%. Since this article, the anchoring and adjustment phenomenon has been widely documented for a variety of measures, such as distance estimates (Kwong & Wong, 2006), price estimates (Mussweiler, Strack, & Pfeffer, 2000), and probability assessments (Plous, 2006).

The manner in which consumers use brand names, however, differs from the typical anchoring paradigm wherein participants are explicitly asked whether the answer to a focal question is smaller or larger than an arbitrary number. According to research on communication norms (Schwarz, 1994), participants in these studies are likely to rely on anchors because they assume that all of the information available to them is relevant. The numbers contained in brand names, however, are not always informative and/or meaningful to consumers. In some cases, these numbers are created for identification purposes only (e.g., Sony’s music player CS200AD). We propose that consumers may associate alphanumeric brand names with certain product-related information, regardless of the original intent behind the brand name. That is, consumers may use alphanumeric brand names as anchors to infer unknown product attributes. For instance, consumers may think that the Airbus A330 has roughly 330 seats, even though the name “A330” has nothing to do with seat capacity.

This anchoring effect should not always occur, however. We propose that at least two boundary conditions moderate this effect. First, consumers should perceive that the numbers contained in alphanumeric brand names are relevant to a given product’s attribute. In the Airbus A330 example, the number “330” should be seen as relevant to the number of seats in the aircraft, but not to its comfort, weight, ticket price, etc. Second, consumers should rely on a heuristic/superficial processing mode rather than a systematic/deliberate one. Previous research (Epley & Gilovich, 2006; Maheswaran, Mackie, & Chaiken, 1992) shows that anchoring is indeed more pronounced when cognitive resources are constrained (e.g., when consumers work under time pressure or are distracted).

Five experiments were conducted to examine our central hypothesis and its two boundary conditions.

3. Experiment 1

The purpose of the first experiment was to demonstrate the basic proposition that consumers rely on alphanumeric brand names to make inferences about product attributes. To this end, we asked
participants to estimate the number of seats in two aircrafts of equivalent capacity (i.e., Airbus A330 vs. Boeing B767).

In two pretests with 66 and 61 undergraduates, we first sought to ascertain whether consumers might in fact relate aircraft brand names to seat capacity. To this end, we asked participants to indicate what the number contained in the brand name Airbus A330 (or Boeing B767) stands for. In the Airbus case, 36 participants (i.e., 55%) associated “330” with the number of seats (e.g., number of passengers, capacity, number of seats). The remaining participants inferred various other meanings such as engine model, length of the aircraft, etc. Similarly, the association made most often in the Boeing case also related to seat capacity. Consistent with our premise, 25 of the 61 participants (i.e., 41%) thought that “767” stands for the number of passengers, capacity, number of seats, etc. The next most common associations regarded speed, code of development, and various other aspects.

Given the apparent connection in the minds of consumers between aircraft names and seat capacity, we expected that the participants in our main study would estimate the number of seats to be greater in the Boeing B767 than in the Airbus A330.

3.1. Method

One hundred forty-five undergraduates who participated in this study for course credit were randomly assigned to either the Airbus or the Boeing condition following a between-subjects design. After being informed that the purpose of the survey was to assess college students’ knowledge about aircrafts, participants were asked how many seats the Airbus A330 (Boeing B767) has. To avoid extreme responses, we collected answers using an 11-point scale, ranging from “1” (100 seats) to “11” (1000 seats) and separated by nine 100-seat increments.

3.2 Results

An analysis of variance revealed that participants did believe that the Boeing B767 (Μ = 463) has more seats than the Airbus A330 (Μ = 395; t (143) = 2.24; p < .05). Consistent with our hypothesis, this finding suggests that consumers do use the numbers contained in alphanumeric brand names as anchors to infer specific product attributes.

4. Experiment 2

The purpose of experiment 2 was twofold. First, for the sake of generalizability, we aimed to replicate our previous findings in a different product domain. Whereas experiment 1 tested our hypothesis in a relatively unfamiliar product category (i.e., aircrafts), experiment 2 employed two famous and familiar soft drink brands as stimuli: Sprite and 7-UP.

Second, we sought to examine a first boundary condition to our effect. We argued earlier that the extent to which anchoring occurs depends on the perceived relevance between a number (e.g., 330) and the product attribute under consideration (e.g., number of seats). To test this idea, we asked participants to infer not one but a variety of attributes from the names Sprite and 7-UP. We hypothesized that participants would associate “7” with the price of 7-UP (note: cans of Sprite and 7-UP typically sell for HK5-10 in Hong Kong, where this experiment was conducted) but not with other attributes, such as volume, vitamin content, etc. This prediction differs notably from what one would expect from the “higher is better” heuristic (Gunasti & Ross, 2010). If the latter was the driving force behind our effect, then participants’ ratings and estimations of all attributes should follow the same rule and, therefore, exhibit the same pattern.

4.1. Method and results

Ninety-six students from HKUST were randomly assigned to either the Sprite or the 7-UP condition before being asked to estimate their drink’s price, volume, history (i.e., launch years), calories, and number of vitamins. Next, participants guessed the average weekly consumption of their drink (i.e., how many cans a person consumes per week) before rating its taste (0 = not at all tasty; 9 = very tasty).

Consistent with our prediction, price estimates in the 7-UP condition (Μ = 6.28) were closer to HK$7 than those in the Sprite condition (Μ = 5.79; t (94) = 2.90, p < .01). Importantly, however, we did not find any significant difference for any other estimates (see Table 1). Hence, these findings provide evidence for the selective nature of our anchoring hypothesis (i.e., that alphanumeric brand names affect only specific consumer inferences). Of note, the “higher is better” heuristic (Gunasti and Ross, 2010) constitutes an improbable alternative explanation for our findings. If the latter underlay our results, one would expect most (if not all) attributes under consideration to follow the same heuristic; hence exhibit the same pattern of results. It was not the case, however.

5. Experiment 3

Because experiments 1 and 2 used real brand names, one may legitimately wonder whether participants’ prior knowledge about these brands may somehow account for (or even contribute to) our findings. To dispel this concern, we opted to feature in the present experiment (as well as in subsequent ones) hypothetical brand names. Furthermore, for the sake of generalizability, we chose to feature in experiment 3 (and in subsequent ones) yet another product category, MP3 players. In sum, experiment 3 aimed to provide more direct evidence for our selective anchoring hypothesis while ruling out alternative explanations rooted either in the specific brand names or in the specific product categories featured in our studies so far.

5.1. Method

5.1.1. Pretest

A review of MP3 brands and models available in Hong Kong (where this study was administered) revealed that such products average HK $500 in price, ranging from HK$100 to HK$1000. Accordingly, we decided to name our MP3 players M-200 and M-900. To verify that participants would perceive these numbers as relevant to price, we asked 32 pretest participants to rate from 1 (very irrelevant) to 7 (very relevant) the relevance of the numbers “200” and “900” for a set of attributes typical of MP3 players (i.e., price, battery life, storage capacity). These attributes were presented in a random sequence to rule out any potential order effect. Consistent with our assumption, participants perceived these two numbers as more relevant to price (Μ = 5.50) than to battery life (Μ = 3.00; t (31) = 7.19, p < .001) or to storage capacity (Μ = 2.97; t (31) = 7.34, p < .001), thereby satisfying our first boundary condition (cf. study 2).

Table 1

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Price</th>
<th>Volume</th>
<th>History</th>
<th>Calories</th>
<th>Vitamins</th>
<th>Storage Capacity</th>
<th>Price</th>
<th>Volume</th>
<th>History</th>
<th>Calories</th>
<th>Vitamins</th>
<th>Storage Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sprite</td>
<td>5.79</td>
<td>353.89</td>
<td>27.15</td>
<td>314.71</td>
<td>2.96</td>
<td>2.81</td>
<td>5.78</td>
<td>0.84</td>
<td>77.05</td>
<td>13.91</td>
<td>330.22</td>
<td>3.14</td>
</tr>
<tr>
<td>7-UP</td>
<td>6.28</td>
<td>368.38</td>
<td>27.49</td>
<td>243.80</td>
<td>2.62</td>
<td>2.53</td>
<td>6.15</td>
<td>0.75</td>
<td>77.42</td>
<td>18.43</td>
<td>182.48</td>
<td>2.49</td>
</tr>
</tbody>
</table>

5.1.2. Predictions

Having confirmed that participants do perceive “200” and “900” to be relevant to price, we predicted that these numbers may then act as implicit Anchors that bias consumers’ perceptions of price. More specifically, we expected that consumers viewing an MP3 player named “M-900” would price the product higher than consumers viewing a similar player named “M-200”. Of note, however, this prediction should hold only as long as consumers do perceive the number-attribute connection. If, for some reason, consumers no longer perceived “200” and “900” to be relevant to price, the anchoring power of the numbers should vanish and consumers’ inferences should not be impacted.

To manipulate the relevance of “200” and “900” to price while keeping the procedure constant across conditions, we devised a task wherein consumers estimated the price of their MP3 player either in HK dollars (i.e., a relevant currency) or in US dollars (i.e., an irrelevant currency). If our theory is correct, anchoring should occur in the HKD condition, but not in its USD counterpart.

5.1.3. Procedure

To test the above predictions, 142 students from HKUST were randomly assigned to one of four conditions following a 2 (brand name: M-200 vs. M-900) × 2 (currency: HKD vs. USD) between-subjects design. After informing participants that the purpose of the study was to investigate how consumers evaluate music players, we presented participants attributes typical of MP3 players, such as brand name, storage capacity, and battery life. This information was kept constant across conditions, with the exception of brand name (i.e., M-200 vs. M-900).

After reviewing this information, participants were asked to estimate the price of their MP3 player. Consistent with our conceptualization, some participants performed this task in a currency that preserved the relevance of “200” and “900” for price (i.e., in HKD), while the rest did not (i.e., in USD).

5.2. Results

A two-way ANOVA was conducted to examine our predictions. As expected given the exchange rate (USD1≈HKD8), the currency of quote yielded a large main effect. Participants in the HKD condition priced their device much higher on average than their counterparts in the USD condition (MHKD=496.72 vs. MUSD=61.38; F (1, 138) = 97.18, p < .001). Neither the brand name main effect (F (1, 138) = 1.68, p > .19) nor the brand name by currency interaction (F (1, 138) = 2.42, p > .12) were significant. Our predictions, however, concerned more specifically the planned contrasts within each currency. To this effect, comparisons between brand names revealed that, when estimates were made in HKD, participants in the M-900 condition priced their device 29% higher on average than their counterparts in the M-200 condition (MHKD=434.39 vs. MUSD=360.83; t (138) = 2.04, p < .05). When estimates were made in USD, however, the two conditions did not differ (MHKD=65.91 vs. MUSD=54.42; t (138) < 1, p > .80).

These findings provide converging evidence for our selective anchoring hypothesis. As long as “200” and “900” appeared relevant for the judgment at hand (i.e., price), the numbers anchored participants’ perceptions about the MP3 players, thereby biasing their estimates. When the connection between numbers and price was absent, however, consumers’ judgments were no longer impacted.

This simple demonstration of the power of alphanumeric brand names should appeal to practitioners (and maybe worry consumer advocates). Naming products is indeed entirely at the discretion of brand managers and is virtually costless. Yet, its potential to influence/ manipulate consumers’ perceptions of a product (e.g., price, value) is so substantial that it is surprising so little of the extant research examines the question.

Of note, the “higher is better” heuristic (Gunasti and Ross, 2010) constitutes again an unlikely alternative explanation for our results. If the latter underlay our findings, one would expect participants to price the M-900 player higher than its M-200 counterpart, regardless of the currency of quote.

6. Experiment 4

Our results from experiments 1–3 suggest that consumers use the numeric component of alphanumeric brands to infer unknown product attributes. For this to occur, however, consumers need to perceive relevance between the number in the brand name and the attribute under consideration (e.g., number of seats, price). Seeking to provide additional evidence for the “relevance” principle of our selective anchoring hypothesis, experiment 4 also aimed to further differentiate our findings from Gunasti and Ross’ (2010). With that in mind, experiment 4 was designed as follows.

6.1. Method

6.1.1. Predictions

We knew from the pretest in experiment 3 that MP3 models in Hong Kong range in price from HK$100 to HK$1000. As such, we hypothesized that an MP3 player with a numeric name outside this range (e.g., 1 or 10,000) should weaken the connection in consumers’ mind between the product’s price and the number embedded in its name. To test this idea, we introduced participants to one of four MP3 players before asking them to estimate the product’s price. The models were identical in all aspects except name (i.e., M-1, M-200, M-900, and M-10000). In line with our conceptualization, we made two predictions. First, as in experiment 3, the numbers “200” and “900” should act as implicit anchors and bias consumers’ price estimates (i.e., M-200>M-900). Second, in contrast, because they fall outside the normal price range of MP3 players, the numbers “1” and “10,000” should not influence consumers’ price estimates (i.e., M-1=M-10000). Of note, the “higher is better” heuristic (Gunasti and Ross, 2010) would again make different predictions under these experimental conditions (i.e., an upward trend in price should emerge across the four conditions).

6.1.2. Procedure

To tease apart the predictions above, 145 undergraduate students from HKUST were randomly assigned to one of four conditions (M-1, M-200, M-900, and M-10000) following a between-subjects design. As in experiment 3, we informed participants that the purpose of the study was to investigate how consumers evaluate music players. Accordingly, participants reviewed attributes typical of MP3 players (i.e., brand name, storage capacity, and battery life) before estimating the price (in HKD) of the product assigned to them. Product descriptions were again held constant across conditions, except for name.

6.2. Results

Replicating our earlier results, planned contrasts revealed that participants in the M-900 condition priced their MP3 player 24% higher than their counterparts in the M-200 condition (M200=489.08 vs. M900=606.82; t (141) = 2.11, p < .05). Because “1” and “10,000” fall outside the normal price range of MP3 players, however, we also hypothesized that participants in the M-1 and M-10000 conditions would not infer price from brand name. Consistent with this prediction, planned-contrast analyses revealed no difference between the two conditions (M1 = 511.44 vs. M10000 = 512.19; t (1) < 1, NS).

As noted earlier, the “higher is better” heuristic (Gunasti and Ross, 2010) would support different predictions. Specifically, the “higher is better” heuristic would forecast a general upward trend across conditions. Yet, none was found (t < 1, NS). Furthermore, price estimates in the M-1 condition were not lower than in the M-200 condition (M1 = 511.44 vs. M200 = 489.08; t < 1, NS). Similarly, price estimates in the M-10000 condition were not higher than in its M-900 counterpart. In
fact, the latter planned-contrast revealed prices marginally lower in the M-10000 condition (M_{900} = 606.82 vs. M_{10000} = 512.19; t (141) = 1.69, p = .09). In sum, experiment 4's results are consistent with our theory and contradict what one would expect if the “higher is better” heuristic were at play (Gunasti & Ross, 2010).

After demonstrating in experiment 1 that alphanumeric brand names can anchor consumer judgments (e.g., number of seats in an aircraft), we examined in experiments 2–4 a first boundary condition to this effect. Specifically, we showed that, for anchoring to occur, consumers should first perceive the numbers embedded in brand names to be relevant to the judgment/attribute under consideration (e.g., price). We demonstrated this principle by either manipulating or manipulating relevance across studies, thereby highlighting the selective nature of consumer anchoring. Along the way, we teased apart the contribution of our work from the merits of Gunasti and Ross’ (2010) “higher is better” heuristic.

In the next study, we shed light on a second boundary condition to our effect, namely, the influence of information processing on consumer anchoring.

7. Experiment 5

As alluded to in our theoretical framework, previous research shows anchoring to be more pronounced when cognitive resources are constrained (e.g., when consumers work under time pressure or distraction) (Epley & Gilovich, 2006; Maheswaran et al., 1992). Therefore, given the heuristic nature of anchoring, we expected that consumers would rely more readily on alphanumeric brand names to infer unknown product attributes when they process information superficially rather than systematically. Experiment 5 was designed to test this idea.

7.1. Method

One hundred ten undergraduates from HKUST were randomly assigned to one of four experimental conditions following a 2 (brand names: M-200 vs. M-900)×2 (information processing: superficial vs. systematic) between-subjects design. As in experiments 3–4, participants were told that the purpose of the study was to investigate how consumers evaluate music players. Accordingly, participants were to review information typical of MP3 players (i.e., brand name, storage capacity, and battery life) and price the product they were assigned (i.e., M-200 vs. M-900). Once again, model descriptions (other than brand names) were held constant across conditions, but the sequence in which participants reviewed information was not. To manipulate information processing, we asked half of the participants to estimate their player’s price before rating its storage capacity and battery life (i.e., superficial processing), whereas the other half did so after (i.e., systematic processing). We predicted that anchoring would be less pronounced in the latter condition where participants processed information more systematically before pricing their product.

7.2. Results

The results of a two-way ANOVA on price estimates revealed a significant main effect of processing style (F (1, 106) = 7.56, p < .05). Specifically, the participants in the pricing-first condition (i.e., superficial processing; M = 705.47) provided higher estimates than their counterparts in the pricing-last condition (i.e., systematic processing; M = 517.86). As predicted, this main effect was qualified by a significant interaction (F (1, 106) = 4.01, p < .05). Anchoring was less likely when participants employed a more systematic (i.e., attribute by attribute) processing strategy (M_{200} = 533.68 vs. M_{900} = 502.04; t < 1) than when they estimated price before analyzing each attribute carefully (M_{200} = 584.57 vs. M_{900} = 826.36; t (106) = 2.73, p < .01). As hypothesized in our second boundary condition, these results support the notion of consumer anchoring being moderated by information processing.

8. General discussion

Consumers often infer unobserved product attributes from available cues (Huber & McCann, 1982; Kardes, Posavac, & Cronley, 2004). In turn, these unobserved attributes have been found to be more influential than readily accessible ones in many consumer decisions (Ford & Smith, 1987). Our research examined across a variety of product categories when and how alphanumeric brand names impact consumers’ inferences of unknown attributes. Building on anchoring theory, we proposed that the numbers contained in alphanumeric brand names might act as implicit anchors, which can subsequently bias consumers’ evaluations of a product’s price, value, weight, capacity, etc. We qualified this proposition, however, by arguing that this anchoring effect is selective; it occurs mostly when (a) the numeric component of a name appears relevant to the judgment at hand and (b) consumers evaluate attributes superficially rather than systematically. Results from five experiments offer converging evidence in support of these hypotheses. Experiment 1 showed that participants’ estimates of seat capacity onboard a Boeing B767 are substantially greater than onboard an Airbus A330, although the actual capacity of these two airplanes is similar. Experiment 2 replicated this finding in the soft drink category by showing that consumers in Hong Kong infer the price of 7-UP to be closer to HK$7 than that of Sprite. More importantly, experiment 2 found this anchoring effect to be selective, i.e., more likely to influence perceptions of product attributes when the numeric component of the name appears relevant to the judgment at hand (e.g., “7” being relevant to price but not to volume, caloric content, etc.). The subsequent two experiments provided more direct evidence for this principle by manipulating the relevance of brand names’ numeric component. When an irrelevant currency was used to infer price (experiment 3) or when the number embedded in brand names fell outside the normal range/distribution of the attribute considered (experiment 4), consumer anchoring disappeared. The fifth and final experiment identified a second boundary condition by demonstrating that consumers are more likely to bias their estimates (e.g., of price) when they process product information superficially rather than systematically.

The present findings contribute to several lines of research. First, as mentioned earlier, previous work on brand naming was largely devoted to the effects of semantic meaning and sound matching of words, paying little attention to the numeric components of brand names. The few papers that broke ground in the investigation of alphanumeric brand names focused primarily on the meaning carried by numbers. For example, Chinese brand names frequently contain “6” or “8” because these are perceived as lucky numbers in Asia (Ang, 1997; Pavia & Costa, 1993). We add to this literature by proposing an additional mechanism through which numeric brand names can influence consumer judgment. Namely, because consumers sometimes infer from numbers contained in brand names certain product attributes, alphanumeric brand names can influence consumers’ inferences and/or distort knowledge retrieval. Compared to previous studies, our theoretical framework is more general and can therefore be applied to a variety of numbers, product classes, and cultures.

Second, our research also adds to the anchoring literature by showing that anchoring can happen spontaneously, without the need for heavy-handed manipulation. As our experiments suggest, consumers can spontaneously use the numbers contained in marketing communications (e.g., advertising, brand names), either to memorize product information and later use them as retrieval cues or to directly infer product attributes. This finding is consistent with existing literature on self-generated anchors. For example, when asked to indicate the height of the second tallest mountain in the world, one might use the height of the well-known Mount Everest and then adjust downward (Epley &
We extend this research stream by examining another type of anchor that, in fact, has no relationship with the focal judgment.

Our third contribution extends recent research on the influence of alphanumeric brand names on consumer inference-making. Most notably, Gunasti and Ross (2010) showed that consumers believe that higher numbers in brand names signal better product configuration or quality. Therefore, when consumers compare side by side two competing products named X-100 and X-200, the latter appears preferable. To complement the “higher is better” heuristic (Gunasti and Ross, 2010), we set out to examine whether consumers might use different heuristics when they cannot easily compare and report products. After all, it is fairly common for consumers to experience the world only one product/service at a time (e.g., in marketing communications, when they receive gifts). As such, what inferences would consumers draw when viewing an advertisement for a single laptop named X-200? Our findings suggest that, under certain conditions, consumers do rely on the numeric component of alphanumeric brands to infer unknown product attributes. Hence, after Gunasti and Ross’ (2010) description of number-based decision-making in joint-evaluation scenarios, our selective anchoring hypothesis offers a clearer understanding of how consumers evaluate a product or service when assessing it on its own.

Lastly, our findings have interesting implications for brand managers and consumers. For example, our results suggest that brand names might not only influence consumer memory and product evaluation but also play an important role in consumer brand knowledge development (Sen, 1999). By strategically naming their products, marketers may be able (and tempted) to seamlessly manipulate consumers’ perceptions. For example, an MP3 player named M-600 sold for $500 should be perceived as offering greater value than one named M-500 sold at the same price. Additionally, according to our findings, the same brand name could lead to different conclusions in different countries. As demonstrated in experiment 2, the “7” in 7-UP is perceived as relevant to product price in Hong Kong where sodas sell for 5–10 HK dollars. It is unlikely, however, that consumers would make the same inference in the United States where soda cans sell for approximately 1 US dollar, or in Japan where they sell for 120 Yen. The results of experiment 3 confirmed this prediction.

Although we identified relevance and information-processing mode as moderators of consumer anchoring, further research is needed to examine other boundary conditions. For example, the likelihood or ease with which a number is connected to a particular product attribute may differ across individuals (e.g., as a function of expertise), leading to different product evaluations. Consider the Canon G10 digital camera, for instance. Novice consumers may infer that this camera has a resolution of 10 megapixels because pixels are the feature most widely used to assess digital cameras. Given that the actual resolution of this camera is 14.7 megapixels, such an inference would be detrimental to the perception of the product and result in less favorable evaluations by novices. Knowledgeable consumers, in contrast, may extract different meanings from this brand name. For example, they may associate “10” with the optical zoom since the optical zoom of most compact digital cameras ranges from 3 to 12. As such, an inferred optical zoom of 10 would lead to a rather favorable impression. Hence, the investigation of expertise as another moderator of alphanumeric brands’ impact on consumer inference might hold promise in the search for a more integrated theory of selective anchoring.

References


“You Lost Me at Hello”: How and when accent-based biases are expressed and suppressed

Ze Wang\textsuperscript{a,1}, Aaron D. Arndt\textsuperscript{b,⁎}, Surendra N. Singh\textsuperscript{c,2}, Monica Biernat\textsuperscript{d,3}, Fan Liu\textsuperscript{a,1}

\begin{flushleft}
\textsuperscript{a} Department of Marketing, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, United States
\textsuperscript{b} College of Business and Public Administration, Old Dominion University, 2147 Constant Hall, Norfolk, VA 23529, United States
\textsuperscript{c} School of Business, University of Kansas, 1300 Sunnyside Avenue, Summerfield Hall room 315-F, Lawrence, KS 66045, United States
\textsuperscript{d} Department of Psychology, University of Kansas, 1415 Jayhawk Blvd. Lawrence, KS 66045, United States
\end{flushleft}

\textbf{A R T I C L E  I N F O}

\begin{flushleft}
Article history:
First received in 11, January 2012 and was under review for 3 months
Available online 23 October 2012
\end{flushleft}

\textbf{Keywords:}
Employee accent
Call service
Justification–suppression model

\textbf{A B S T R A C T}

This research examines customer biases relating to employee accents in call service encounters. Extant research and practitioners generally assume that customers automatically evaluate call service employees with a nonstandard accent lower than employees with a standard accent. However, using the justification–suppression model as a framework, we argue that customers frequently suppress accent biases toward call service employees. We conduct three empirical studies, and our findings indicate that customers rate employees with an accent receiving a negative bias lower only when a service outcome is unfavorable for customers. In contrast, accents receiving a positive bias only impact customer evaluations when service outcomes are favorable for customers. Additionally, we demonstrate that the suppression and justification of accent biases rely on both cognitive and affective mechanisms. Finally, we show that customers who are informed of the frequency of a favorable vs. unfavorable outcome are more likely to suppress biases.

Published by Elsevier B.V.

\textbf{1. Introduction}

"Are you calling from India?"
"No, I'm calling from Modesto, California."
"Well, you sound Indian."
"I've only been here for two months and haven't got the accent right."

A conversation between a potential U.S. customer and a 22-year-old Indian call center worker (Nadeem, 2011).

"I remember the first time it happened. I had just started. I introduced myself, and the customer immediately demanded to know where I was calling from. I told him India and he just said 'F*** off you job-stealing Paki' and slammed the phone down. It left a really sour taste to hear that from an Englishman. I was shaking with anger" (Foster, 2005).

This experience was told by Ian Hussey, a young British student who briefly worked as an intern at an Indian call center, calling British customers to switch telecom services — reported in the Telegraph.

While there have been a few highly publicized cases of companies (e.g., Apple, CapitalOne, Dell, Delta) closing their foreign call centers due to customer dissatisfaction and moving operations back to the U.S., the trend points to an increasing overseas movement of service call centers owing to the compelling cost savings. Accordingly, there is a need to understand U.S. consumer sentiment toward foreign call service employees (Thelen, Yoo, & Magnini, 2011). While extant research has found that customers evaluate call services differently based on call center location and cultural similarity (Bharadwaj & Roggeveen, 2008; Roggeveen, Bharadwaj, & Hoyer, 2007), consumers usually do not know where a call center is located and, even if they ask, employees are sometimes instructed to lie (Poster, 2007). Consistent with Thelen et al. (2011), we opine that consumers use the employee’s accent to surmise employee nationality and call service characteristics. However, to our knowledge, there are no rigorous studies explaining how U.S. customers use accent as a feature when evaluating call service employees. Based on research in linguistics and social psychology, we address this research gap by examining U.S. customer biases toward foreign-accented employees in a call (i.e., voice only) service environment.

Accented speech, especially in the absence of visual cues (as is the case in call communications), provides two pieces of information for the listener: (1) linguistic, relating to the content of the speech, and (2) indexical, relating to inferences about the speaker (Levi & Pisoni, 2007). With respect to content, the main issue is intelligibility: can the customer understand what the employee is trying to communicate? While unintelligibility could be a major source of frustration and customer dissatisfaction, we argue that accent-revealed indexical information is also a major issue. Research shows that listeners can still hold
negative biases toward nonnative accents even when the content of the speech is entirely understood (Hosoda & Stone-Romero, 2010). A nonnative speaker’s accent triggers categorization in a prompt, automatic, and occasionally unconscious manner (Rakić, Steffens, & Mummendey, 2011). Listeners can identify an accented speaker’s ethnic or cultural group membership by just listening to 30 ms of speech (Flege, 1984) or as soon as the speaker says, “hello” (Baugh, 2000). Indeed, even when listeners do not recognize a speaker’s specific accent (e.g., Korean), they still tend to make snap judgments about the speaker (Lindemann, 2003). Individuals with a nonstandard accent or dialect may be perceived as less competent, less intelligent, and less industrious (Edwards, 1982).

Several factors have contributed to the increased negative stereotyping of foreign-accented employees in call service encounters, including the bleak U.S. economic climate, with high unemployment rates, heightened media attention to outsourcing as a tool for exporting jobs overseas, and negative publicity about foreign-accented service employees’ competence and professionalism (Modic, 2007; Sandberg, 2007). Two things are noteworthy, however. First, not all accents are created equal. In the U.S., people who speak with a standard British accent (often called the ‘Queen’s English’) or with a French accent may be perceived as sophisticated, cosmopolitan, and well-educated (Cargile, 2000; Stewart, Ryan, & Giles, 1985). Asian-accented English speakers, on the other hand, are perceived less favorably, considered to be poorer communicators and less effective (Hosoda & Stone-Romero, 2010). Second, it is not inevitable that these biases will alter customer attitudes toward the employee and the employee’s service abilities. According to the justification-suppression model (JSM), prejudice toward a group or individual may be suppressed for any number of reasons, including ideologies promoting liberalism, humanitarianism, and social equality; the desire to maintain a non-prejudiced self-image; and social norms promoting anti-prejudice or “political correctness” (Crandall & Eshleman, 2003). Stereotypes are also less likely to be applied when individual group members do not “fit” stereotypical perceptions (Breuer, 1988). For example, King, Shapiro, Hebl, Singletary, and Turner (2006) found that obese people who wore attire inconsistent with stereotypes of obesity were less likely to be negatively stereotyped.

Given the number of employees with nonnative accents providing call service to U.S. customers, it is important to understand customers’ accent-induced prejudices. To this end, the purpose of our research is to examine how accent biases influence customer evaluations of call service interactions. Consistent with the JSM model, we propose that customers tend to suppress overt negative biases toward an accented employee if the outcome of the service encounter is favorable for the customer. However, if the outcome of a service encounter is unfavorable, customers will be less likely to suppress biases because they will have less emotional control and because they will use the negative accent bias as an explanation for the unfavorable service. In contrast, we argue that positive biases will manifest when outcomes are favorable because a positive bias is consistent with a favorable outcome, and people generally do not feel the need to suppress positive biases. Finally, we contend that providing information about the likelihood of a particular outcome should reduce the interaction between accent biases and outcome. When customers know the probability of obtaining a favorable outcome, they are less likely to relate a particular outcome to their accent biases.

We begin with an overview of relevant literature in marketing, sociolinguistics, and social psychology on stereotyping and accent prejudice. Next, we present hypotheses regarding the moderating role of service outcome on the relationship between negative accent biases and customer evaluations of call service employees. These hypotheses are then tested in Study 1 using both a lab and a field test. In Study 2, we develop and test hypotheses regarding the cognitive and affective mechanisms underlying accent biases. Study 3 extends the first two studies by developing and testing hypotheses pertaining to positive accent biases (standard British accent) and negative accent biases (Indian accent). Study 3 further discusses and tests the impact of accent-based biases when customers have more information about typical service outcomes. We conclude with a discussion of the main findings, including their theoretical and managerial implications.

2. Conceptual background

2.1. Accent as a basis for stereotypical judgments

In the context of voice-to-voice service encounters, employee accent is likely to play a crucial role in stereotypical inferences. Accent is defined as “a manner of pronunciation with other linguistic levels of analysis (grammatical, syntactical, morphological, and lexical) more or less comparable with the standard language” (Gluszek & Dovidio, 2010, 215). A common way to categorize accent relies on its standardness or “the systematization and acceptance of a formal set of norms defining correct usage among language users” (Morales, Scott, & Yorkston, 2012). In the U.S., the accent considered to be relatively free of regional influence is referred to as the ‘standard American accent’ (Ghorshi, Vaseghi, & Yan, 2008).

Prior research has shown that a nonstandard accent can be disadvantageous (Cargile & Giles, 1997). Frequently, there are negative biases against nonstandard accents, particularly those associated with disadvantaged and low-prestige minority groups, and listeners may display irritation or other forms of prejudice toward a speaker with such an accent (Hosoda & Stone-Romero, 2010). For example, in the context of criminal investigations, suspects with regional, nonstandard accents appeared significantly guiltier or “typically criminal” to respondents compared with standard-accented counterparts (e.g., Dixon & Mahoney, 2004). Similarly, in employment decisions, Spanish-, Asian-, and African-accented speakers – particularly those with heavy accents – receive lower employability ratings than their counterparts with a standard American accent (Purkiss et al., 2006).

These accent-based inferences, in turn, affect marketing outcomes. Among U.S. respondents, Mexican- or Greek-accented spokespersons and salespeople were perceived as less intelligent, honest, credible, and professional than their standard American-accented counterparts, thus leading to lower purchase intentions among buyers (DeShields & de los Santos, 2000). Expanding this line of research, we suggest the possible existence of similar accent-based biases in call service encounters. Talking to a nonnative-accented employee may evoke a customer’s negative predispositions about that employee’s competence and professionalism, heighten the customer’s dislike, and/or increase the customer’s annoyance. These negative biases could be exaggerated further by direct cultural learning or intergroup conflict. Mass media and various consumer-generated contents encourage accent stereotypes and often emphasize negative assessments of foreign-accented service employees (Modic, 2007; Sandberg, 2007). The tactic of outsourcing jobs to foreign countries also has received negative publicity, such that many people believe that outsourcing reduces the number of U.S. jobs available and hurts national interests.

Even when listeners have little linguistic expertise or knowledge, they can make basic distinctions among broad accent categories (e.g., Spanish-accented vs. British-accented English), although they generally cannot distinguish between more nuanced categories, such as varieties of Spanish-accented English (e.g., Cuban, Costa Rican, Argentinean, Puerto Rican) (Podberesky, Deluty, & Feldstein, 1990). As previously mentioned, Lindemann (2003) found that, even when listeners do not recognize a speaker’s specific accent, listeners still make stereotypical inferences about the speaker. Therefore, our intent is not to provide a comprehensive catalog of accent biases or to investigate the accuracy of customers’ accent perceptions but rather to determine whether customers’ accent-based biases are suppressed or reinforced in certain conditions.
2.2. Suppression and expression of accent-based biases

Building on recent research in social cognition, we propose that biases resulting from accent-based stereotypes may be more complex than previously believed. Although stereotypes are recalled reflexively, people do not automatically act on their prejudices (Lowery, Hardin, & Sinclair, 2001). According to the JSM (Crandall & Eshleman, 2003), people instead confront two competing motivations: (1) the automatic impulse to apply quick stereotypical evaluations and (2) the need to suppress any evaluations that appear inappropriate (Lowery et al., 2001). Stereotypical evaluations, particularly negative ones, are often suppressed (Crandall & Eshleman, 2003). Modern U.S. culture indicates a strong dislike for unsuppressed prejudice and for those who exhibit it (Mae & Carlson, 2005), who are often described using negative terms such as “bigot” or “chauvinist.” Thus, many people work to suppress their prejudices (King et al., 2006).

The initial uncontrolled prejudice that people feel is called “genuine” prejudice, which is then either suppressed or justified. However, if new information appears that is consistent with the negative bias, the individual may switch from suppressing to justifying the prejudice; hence, prejudice is a dynamic process (Hegarty & Golden, 2008). When people are able to justify a genuine prejudice that had been previously suppressed, it causes a sense of psychological “relief.” This urge for relief motivates people suppressing prejudice to look for information that is consistent with the negative bias (Crandall & Eshleman, 2003).

Just as customers are subject to the processes that lead to accent-linked biases, they are also subject to social norms about prejudice or stereotypical judgments. It is the norm in the U.S. to suppress most types of prejudice (King et al., 2006); hence, accent prejudices are also likely to be suppressed. However, we argue that information from the service encounter, which is consistent with a customer’s genuine prejudice, can be used by customers to justify their prejudice, in which case the accent prejudice will be expressed. Arguably, the most salient information about the service encounter is the degree to which the service outcome is favorable or unfavorable for the customer. If the service outcome is favorable for the customer, it is not consistent with a negative bias, and thus, it cannot be used to justify an accent prejudice. Accordingly, the prejudice will remain suppressed. However, when the outcome of a call service encounter is not favorable, it is consistent with a negative bias, and the customer can use the service outcome to justify their prejudice. In that case, the accent prejudice will be expressed and can then be used to make attributions about the cause of the service failure. This argument is consistent with research by Sinclair and Kunda (2000), who find that women are given lower teaching evaluations than men, but only when students receive lower grades. Formally:

**H1.** Customers will rate employees with an accent receiving a negative bias lower than employees with a standard accent only when a service outcome is unfavorable for customers.

2.3. Service attributions as assessments of the suppression-expression process

After a service encounter, customers feel satisfaction or dissatisfaction with the outcome (e.g., Bitner, 1990; Cowley, 2005), and people try to find explanations for the success or failure of an experience, especially if the encounter is important to them (Taggar & Neubert, 2004). Consistent with the JSM, people find relief when they are able to justify genuine prejudice. Accordingly, we expect that people will feel inclined to elaborate more on the employee’s contribution to the service encounter when they are able to justify genuine prejudice. Therefore, when service outcomes are unfavorable for customers, we argue that customers will provide longer attributions about the employee when the employee has an accent receiving a negative bias compared to when the employee has a standard accent. This bias in customers’ attributions should not be observed when the service outcome is favorable because that scenario leads to the suppression of prejudice. Formally,

**H2.** Customers will describe the employee contribution to an unfavorable service outcome in more detail when the employee has an accent receiving a negative bias compared to when the employee has a standard accent.

3. Study 1

Study 1 investigates how consumers’ judgments of a service encounter may be biased by employees’ accents. As previously discussed, we expect a negative bias toward call service employees with Indian accents (e.g., Modic, 2007), and accordingly, we will compare the Indian accent with the standard American accent. Study 1 consists of a lab test that will examine H1 and H2 and a field test, where we will further investigate H1.

3.1. Lab test

Participants were asked to listen to a recorded phone conversation in which a customer calls into a bank to update address information. The phone conversations were designed by the researchers based upon real cases described on a consumer website about banking services. The outcome of the service was manipulated to be either unfavorable or favorable. We define the service outcome according to whether customer requests are granted (i.e., the customer receives what he or she is asking for). For example, in the unfavorable condition, the customer could not update address information over the telephone and was required to make an inconvenient trip to a local branch. In the favorable condition, the customer smoothly updated the address information during the service call.

Generally, classic sociolinguistic research on attitudes toward accents (e.g., Giles & Powesland, 1975) has employed a matched-guise technique in which the same speaker imitates different accents. However, more recent research has recommended using native speakers to enhance reliability (Stockwell, 2002). Consistent with that recommendation, we employed two female voice actors with genuine accents to play the part of the employee, one with an Indian accent and one with a standard American accent. The customer was voiced by a male actor with a genuine standard American accent and remained the same in all conditions. The actors were asked to maintain a neutral voice (neither happy nor sad) during the conversation. Different accent versions were carefully matched in terms of the pacing of the conversation, speech intonations, pitch, and intensity. We analyzed these voice qualities using Praat freeware (available at www.fon.hum.uva.nl/praat/).

3.1.1. Design, participants, and procedure

This study used a 2 (accent: standard American vs. Indian) × 2 (service outcome: favorable vs. unfavorable) between-subjects design. One hundred twenty-two undergraduate students at a large U.S. Midwestern university participated in the study in exchange for extra credit. Upon their arrival at a computer lab, the participants were randomly assigned to one of four experimental conditions. As a cover story, participants were told that we were interested in understanding how to improve employees’ quality of service over the phone. Similar to the role-playing technique used by Bitner (1990), participants were asked to listen to a recorded customer service call in which they were instructed to take the perspective of the customer. Because service duration has been shown to influence customer evaluations (Yeung & Soman, 2007), all recordings were of comparable length (between 2 min 30 s and 2 min 45 s). Finally, participants completed a post-service questionnaire that measured the dependent variables.
3.1.3. Manipulation check

The manipulation of employees’ accents was successful. Participants rated employees’ speech, anchored at 1 = “very accented” and 7 = “unaccented (native-like).” The Indian-accented employee was perceived to have a significantly stronger accent than the American-accented employee (M_Indian = 4.45, M_American = 5.37; F(1, 118) = 116.81, p < .01). No other effects were significant, although the Indian-accented employee received lower ratings with respect to understandability (M_Indian = 6.23, M_American = 4.57; F(1, 118) = 4.45, p < .01). However, all employees’ understandability ratings were above the midpoint of the scale, and there were no significant differences in participants’ comprehension of the conversation.

3.1.4. Lab test results

For customer perceptions of rapport, the full-factorial 2 (accent) × 2 (service outcome) analysis of variance (ANOVA) revealed the predicted two-way interaction (F(1, 118) = 4.45, p < .02), shown in Fig. 1. With an unfavorable service outcome, participants reported significantly lower rapport with the Indian-accented employee than with the American-accented employee (M_Indian = 3.00, M_American = 3.91; F(1, 118) = 4.89, p < .03). However, when the service outcome was favorable, customer–employee rapport ratings did not differ across the Indian and American accent conditions (M_Indian = 4.16, M_American = 3.90; F(1, 118) = .47, NS). These findings support H_2.

Next, we examined post-service attributions. A 2 (service employee’s accent) × 2 (service outcome) ANOVA showed a significant two-way interaction (F(1, 118) = 5.71, p < .02), as shown in Table 1. Contrast analyses further revealed that, in support of H_3, when the service was unfavorable, respondents in the Indian accent condition described the employee contribution in more detail compared with the respondents in the American accent condition (M_Indian = 23.13, M_American = 13.19; F(1, 118) = 16.81, p < .01). However, when a service outcome was favorable, employee accent had no impact on the description of employee contribution (M_Indian = 21.24, M_American = 21.79; F(1, 118) = .21, NS).

3.2. Field test

In our experimental design, only a limited amount of information was provided to respondents, thus allowing us to ensure the internal validity of our manipulations. For instance, we used a fictitious bank in the scenarios so that respondents would not know the firm’s history or brand image. However, in real-life settings, customers are generally familiar with the firms they call through branding and previous

Table 1

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Favorable service outcome</th>
<th>Unfavorable service outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>American</td>
<td>Non-Western</td>
</tr>
<tr>
<td>Study 1: lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer–employee rapport</td>
<td>4.01</td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>Elaboration on employees’ performance</td>
<td>21.79</td>
<td>21.24</td>
</tr>
<tr>
<td>Total number of words</td>
<td>(11.77)</td>
<td>(12.11)</td>
</tr>
<tr>
<td>Study 1: field*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer–employee rapport</td>
<td>5.75</td>
<td>5.64</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>Customer satisfaction</td>
<td>5.78</td>
<td>5.95</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Study 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer–employee rapport</td>
<td>4.02</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>(1.28)</td>
<td>(1.58)</td>
</tr>
<tr>
<td>Customer satisfaction</td>
<td>4.34</td>
<td>4.87</td>
</tr>
<tr>
<td></td>
<td>(1.41)</td>
<td>(1.56)</td>
</tr>
<tr>
<td>Future behavioral intentions</td>
<td>4.54</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>(1.31)</td>
<td>(1.09)</td>
</tr>
<tr>
<td>Changes in customers’ mood</td>
<td>5.01</td>
<td>5.22</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Employee performance</td>
<td>5.29</td>
<td>5.33</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.91)</td>
</tr>
</tbody>
</table>

Standard deviations are in parentheses.

* Controls include respondent age, respondent gender, employee gender, and whether the call involved at least one transferred.
experience. Additional information could influence how much attention respondents devote to accent and service outcome. Accordingly, to ensure that the interaction between service outcome and accent applies in real call service encounters, we retested H1 using survey data from actual customer experiences with call–in service interactions.

Additionally, our previous experiments were tested using a student sample. It is possible that students are more likely to repress negative biases toward accents because they are exposed to international students and are less likely to have been in a job threatened by foreign outsourcing. Therefore, it is important to survey a more diverse sample in terms of age and background. Furthermore, our experimental scenarios were set in a single industry, banking. As customers may have different attitudes toward foreign accents in other industries, this study will examine customer experiences across a broad range of industries. Finally, our lab test measured only customer–employee rapport, a direct assessment of the employee. However, service providers have a direct and profound influence on the service experience itself (Bitner, 1990). For example, Bitner (1990, 69) explains that, in many cases, “discrete encounters are the service from the customer’s point of view” (italics added). Hence, it is possible that customer biases directed toward employees also impact overall customer satisfaction with the transaction. Accordingly, the field test will measure both customer–employee rapport and customer satisfaction with the transaction, defined as “the cognitive assessment of a customer’s emotional experience” as it relates to a single interaction (Hennig-Thurau et al., 2006, 60).

3.2.1. Design, participants, and procedure

For this study, we used an online survey about a recent call service experience. The sampling frame included U.S. participants of a large crowdsourcing marketplace, Mechanical Turk. A number of researchers have supported the efficacy of using MTurk respondents (Buhrmester, Kwang, & Gosling, 2011; Goodman, Clyde, & Cheema, in press). For example, Goodman et al. (in press, p. 1) explain that “MTurk offers a highly valuable opportunity for data collection.” Respondents were screened to ensure that they were from the U.S., spoke English as their first language, had called for customer service within the last month, and had maintained sufficient work quality on previous MTurk assignments. Respondents were paid $0.50 cents to complete the survey. A total of 595 qualified respondents completed the survey.

Qualified respondents were asked to recall a phone call made to customer service within the last month. Service outcome was then measured by asking respondents to describe the result of the service encounter in an open-ended response and by directly asking respondents whether they felt the service outcome was favorable or unfavorable. As accent is usually the only information customers have regarding the location of a call service employee, employee accent was measured by asking respondents to guess as to whether the employee was located in the U.S. or in a foreign country, and if in a foreign country, which country. Of the 91 calls where respondents thought the employee was located in a foreign country, the majority guessed an Asian country. Only 4 guessed the employee was from a Western country, and those responses were excluded. As such, this study compares the standard American accent with non-Western accents. Early analysis indicated that considerably more respondents believed that the call service employee they spoke to was located in the U.S., particularly with respect to calls that were rated as having favorable outcomes. Littell, Stroup, and Freund (2002) suggest that sample size imbalance when comparing groups results in unstable findings. Accordingly, we randomly selected 25% of the call service employees located in the U.S. to reduce the imbalance.

To decrease bias between satisfaction with the call and the service outcome, two coders independently categorized the open-ended responses about the service outcome as either favorable or unfavorable. Ninety-seven respondents were dropped due to unclear open-ended answers. The inter-rater agreement for the remaining respondents was 81.7%, and differences were resolved by a third independent coder. Next, we assessed the agreement between the respondent’s evaluation of the outcome and the coder’s evaluation. The agreement between the respondents and the coders was approximately 96%. For hypothesis testing, we used the coder’s assessment of service outcome. The final sample size was 197. The breakdown by outcome and accent is shown in Table 2.

3.2.2. Dependent measures

H1 was tested using two dependent variables, customer perceptions of rapport with the employee and customer satisfaction. Consistent with the lab test, customer–employee rapport was measured using a 4-item, 7-point Likert scale from Hennig-Thurau et al. (2006). Customer satisfaction was also measured using a 4-item, 7-point Likert scale from Hennig-Thurau et al. (2006). Control variables included employee gender, respondent gender, respondent age, and whether the phone call included a transfer, as a transfer could impact recollection about the service employee who was primarily responsible for assisting the customer.

3.2.3. Field test results

The hypothesis was tested using ANCOVA. For the model used to test rapport, the interaction between service outcome and accent was significant (F(1,196) = 5.36, p < .05). When the outcome was unfavorable for the customer, the perception of rapport was significantly higher for employees with an American accent than for employees with a non-Western accent (American = 5.29 vs. non-Western = 1.73, F = 11.47, p < .001). However, when the outcome was favorable for the customer, the perception of rapport was not significantly different (American = 5.75 vs. non-Western = 5.64, F = 0.00, ns). The interaction means are shown in Table 1 and Fig. 2.

Similarly, for the customer satisfaction model, the interaction between service outcome and accent was significant (F(1,196) = 7.40, p < .01). When the outcome was unfavorable for the customer, customer satisfaction was again, significantly higher for employees with an American accent than for those with a non-Western accent (USA = 2.16 vs. non-Western = 1.34, F = 9.12, p < .01); however, when the outcome was favorable for the customer, customer satisfaction was not significantly different (American = 5.78 vs. non-Western = 5.95, F = 0.92, ns). The findings for both rapport and customer satisfaction support H1. Hence, evidence suggests that the interaction between accent and service outcome holds in real call service environments.

3.3. Post-test of accent bias

Using both a student sample and a broader online sample, we found that employees with non-Western accents were only rated lower than employees with standard American accents in the unfavorable outcome condition, but not in the favorable outcome condition. However, it is unclear whether respondents were truly suppressing a pre-existing bias or whether the bias manifested as an explanation for the service failure. Accordingly, we conducted a post-test with 121 students and 75 U.S. respondents from MTurk to establish whether the biases found in Study 1 were pre-existing. Consistent with the procedure used by Morales, Scott, and Yorkston (2012), we used a between-subjects design in which respondents listened to five recorded statements spoken by either a standard American-accented speaker or an Indian-accented speaker. After
each statement, respondents were asked to rate the likability of the speaker on a 7-point Likert scale. Likability ratings for the five statements were averaged together to create a single measure of likability. Controls, including respondent age, respondent gender, and respondent pre-study mood, were measured using a 4-item, 7-point Likert scale adapted from Tsai and Huang (2002) and Hennig-Thurau et al. (2006).

An ANOVA test was used to analyze the results. As shown in Table 3, the American accent was rated as significantly more likable than the Indian accent by students ($F(1, 120)=41.14, p<.001$) and by MTurk respondents ($F(1, 74)=15.89, p<.001$). This finding suggests that there is a prevalent pre-existing bias in the U.S. against speakers with an Indian accent.

### 3.4. Discussion

Consistent with the findings from our post-test, it is commonly believed that nonstandard accents (e.g., Indian) will always be received less favorably than standard accents (e.g., standard American). As such, managers have adopted a number of costly measures to reduce accent biases such as adopting speech training programs or moving customer service centers out of countries subjected to negative accent biases. On the basis of such beliefs, managers may be tempted to reject employees with Indian accents, even in circumstances without accent-linked biases. However, the results of Study 1, demonstrated in both a lab test and a field test, indicate that negative biases against employee accents only reduce employee ratings when a service outcome is unfavorable for customers. Furthermore, the lab study shows that respondents only offer longer explanations about employee contribution to the service outcome when the service outcome is unfavorable. Thus, U.S. customers appear to suppress their accent biases when service outcomes are favorable.

However, while Study 1 demonstrates when customers express negative biases against accented call service employees, it does not explain the underlying mental mechanisms for these biases. According to Crandall and Eshleman (2003), suppressing prejudice requires both cognitive and affective effort. It is not clear from Study 1 whether unfavorable service outcomes are used cognitively as an excuse by customers to justify prejudice or whether unfavorable service outcomes cause customers to lose emotional control such that they are no longer able to suppress prejudice. Therefore, in Study 2 we investigate the role of cognition and mood in the suppression/expression of accent prejudice.

### 4. Study 2

In Study 2, we examine the role of cognition and affect in accent prejudice. Whereas the mental processes used to justify biases are cognitive, the mechanism underlying suppression requires both affective and cognitive effort (Crandall & Eshleman, 2003). Crandall and Eshleman (2003, 422) explain that suppressing prejudiced thoughts, emotions, and feelings requires ongoing “mental energy” and leads to a negative feeling. When service outcomes are unfavorable for customers, we expect that they will frequently become upset and agitated. People who are in highly emotional states are less able to maintain control over their biases and prejudices (Crandall & Eshleman, 2003). Hence, when a service encounter has an unfavorable outcome, it is possible that customers may no longer be able to suppress their prejudices. As customers become increasingly less able to suppress the accent prejudice, the stereotypical inferences provide the most accessible source of information to use in service attributions. Therefore, we contend that service failure may bring forth prejudice using two pathways, 1) the cognitive pathway, which provides justification for prejudice, and 2) the affective pathway, which makes it more difficult for customers to continue suppressing prejudice. Taken together, we expect that this suppression and justification process will reduce or release the private experience of prejudice (affective responses and cognitive inferences) and public expression (e.g., reported judgment of the service encounter). Formally:

![Fig. 2. Rapport and customer satisfaction as a function of service outcome and employee’s accent.](image)
H3. The outcome-moderated accent stereotyping effect is mediated by both affective (changes in customers’ mood) and cognitive (customers’ cognitive inferences) mechanisms.

4.1. Design, participants and procedures

Study 2 used a 2 (service employee’s accent: standard American vs. Indian) × 2 (service outcome: favorable vs. unfavorable) between-subjects design. Sixty-three undergraduates at a large U.S. Midwestern university participated in this experiment in exchange for extra credit. Participants were randomly assigned to one of four conditions and listened to a taped telephone call between a customer and a bank employee, from the perspective of the customer. After listening to the call, participants completed a questionnaire that was inclusive of the dependent measures and control variables.

Moreover, we adopted a mood-neutralizing procedure to ensure that participants’ pre-encounter moods were comparable. With this treatment, unlike a simple measurement of pre-encounter mood, we could rule out the confounding effect of pre-encounter affective states on the accent stereotyping effect, such as their ability to increase or decrease a person’s propensity to use stereotypical heuristics (Park & Banaji, 2000). Immediately after the service encounter, we measured post-service mood. Finally, we asked participants to rate employee performance on a multi-item scale.

4.2. Measures

The dependent measures included an evaluation of employee performance, customer satisfaction, and change in the respondent’s emotional state. Evaluation of employee performance was measured using a 7-item scale (α = .87), adapted from Cronin, Brady, and Hult (2000), that focuses on perceptions of employees’ competence, trustworthiness, friendliness, and understandability. Customer satisfaction relied on a 4-item scale, adapted from Hennig-Thurau et al. (2006).

We used a two-step procedure to measure the changes in the customers’ emotional states. First, before listening to the service call, participants underwent an adapted version of Velten’s (1968) mood induction procedure – reading 50 neutral sentences (e.g., scientific facts), each displayed on a screen for 10 s – to neutralize their pre-encounter moods. Participant pre-encounter moods were measured on a 4-item, 7-point scale (sad/happy, bad mood/good mood, irritated/pleased, depressed/cheerful; r = .77) and did not differ significantly across experimental conditions (F(1, 59) = 3.32, NS). Second, after listening to the service call, emotional state was measured using a 6-item scale (warm-hearted, pleased, cheerful, upset, angry, annoyed), adapted from Tsai and Huang (2002) and Hennig-Thurau et al. (2006). A change in the emotional state represents a difference between the respondent’s pre- and post-call mood. Note that we strategically chose different scale items for the pre- and post-service mood measures to avoid raising participants’ suspicion about the association between the two parts of the study.

As a manipulation check of employee accent, we asked participants to guess the locality of the service provider (i.e., somewhere in the participant’s state, somewhere in the U.S., or a foreign country).

4.3. Results

4.3.1. Customers’ evaluations of the service encounter

We created an index for each customer outcome variable: customer–employee rapport, customer satisfaction, and future behavioral intentions. We provide the means and standard errors of these dependent variables in Table 1. Because the results of the 2 (employee accent) × 2 (service outcome) ANOVAs were strikingly similar, we report only one set of ANOVA test results: those of the customer–employee rapport index. The predicted interaction of employee accent and service outcome (F(1, 59) = 3.96, p = .05) arose, such that when the service outcome was favorable, employee accent had no effect (M_A = 4.50, M_I = 4.02; F(1, 59) = 2.34, NS), but when the service outcome was unfavorable, customer satisfaction ratings fell significantly in the Indian accent condition relative to those in the standard American condition (M_A = 3.86, M_I = 3.09; F = 6.04, p < .05, one-tailed).

4.3.2. Multi-mediation test

For this analysis, we conducted a multi-mediation test because it allowed us to investigate more than one mediator simultaneously. Specifically, we examined the extent to which accent→service outcome→service evaluation mediated by both affective mechanisms (e.g., changes in customers’ positive and negative feelings) and cognitive mechanisms (e.g., customers’ inferences about the employees’ competence and reliability). If either affect or cognition fully mediates the model by itself, then the finding would suggest that one mechanism plays a greater role in accent prejudice. However, if neither cognition nor affect fully mediates the model by itself, but together they do, then we find that a service failure elicits prejudice using both cognitive and affective pathways.

Following Muller, Judd, and Yzerbyt (2005) and Preacher and Hayes (2008), we derived a mediated moderation model with customer mood and employee performance as two potential mediators (Fig. 3). These procedures generated a 95% confidence interval around the indirect effect, and mediation exists if zero falls outside of that confidence interval. Bootstrapping procedures (Preacher, Rucker, & Hayes, 2007) for multiple mediator models (Preacher & Hayes, 2008) are appropriate for our study for three reasons. First, when working with small samples, bootstrapping overcomes the traditional constraint of assuming a normal distribution (MacKinnon, Lockwood, & Williams, 2004). Second, as all proposed mediators can be tested simultaneously, a common problem for mediation tests in which an omitted mediator could lead to a biased parameter estimate is not an issue (Judd & Kenny, 1981). Third, the specific indirect effect of each mediator can be tested while controlling for all other variables in the method, and specific indirect effects can be compared for mediation strength. Thus, the analysis and bootstrap estimates were based on 5000 bootstrap samples.

Table 4

<table>
<thead>
<tr>
<th>Transparent information context</th>
<th>Ambiguous information context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable outcome</td>
<td>Unfavorable outcome</td>
</tr>
<tr>
<td>American</td>
<td>British</td>
</tr>
<tr>
<td>Customer satisfaction</td>
<td>5.04 (1.20)</td>
</tr>
<tr>
<td>Percentage attribution to dispositional factors</td>
<td>50.00</td>
</tr>
</tbody>
</table>
The total effect of accent × outcome on customer–employee rapport was significant (β = −1.25, t(63) = −1.99, p = .05), whereas the direct effect was not (β = .10, t(63) = .24, NS). The total indirect effect through two mediators was significant, with a point estimate of −1.35 and a 95% bias-corrected and accelerated bootstrap confidence interval (CI) ranging between −2.48 and −.34. The proposed mood and cognitive mechanisms fully mediated the association between accent × outcome and customer service evaluations. The specific indirect effects for each proposed mediator showed that changes in customers’ feelings and employee performance were unique mediators, with point estimates of −.79 and −.56 and 95% CIs of −1.60 to −0.15 and −1.15 to −.06, respectively. A contrast test between the significant indirect effects revealed that the specific indirect effects of customer mood and employee performance were not significantly different (95% CI = −1.04 to .37). The estimates for each specific path, which we report in Fig. 3, thus support H2.

4.4. Discussion

Study 2 supports our assertion that both cognitive and affective mechanisms play a role in the suppression/justification of accent biases. When a service outcome is unfavorable, customer interactions with nonstandard-accented employees lead to changes in the customer’s emotional state (i.e., feeling less pleasant and more annoyed) and expressions of stereotypical beliefs (i.e., Indian-accented employees are not as competent). Both mechanisms are unique mediators and not significantly different in terms of mediation strength. Additionally, in Study 2, we replicate the findings of Study 1 using multiple measures. Accent biases influence not only customers’ judgments of interpersonal interactions but also their overall satisfaction with the encounter and their future loyalty intentions.

5. Study 3

To test the boundary conditions of the proposed outcome-moderated accent-linked biases, in Study 3, we made three extensions. First, we examined both positive (British) and negative (Indian) accent biases and compared them with the standard in-group condition (standard American). Second, we explored whether informing customers about the frequency of a favorable outcome for this type of service call would cause them to be less likely to rely on accent stereotypes. Third, we increased the generalizability of the previous findings by considering another service context, also described on a consumer website about banking service, in which the customer requested an overdraft fee cancellation — a more complex service transaction than simple address updating. In the favorable outcome condition, most of the overdraft fee was reimbursed ($66 of $99); in the unfavorable outcome condition, none of the overdraft charges were refunded.

5.1. Positive accent biases

In Study 1 and Study 2, we consistently found that a negative accent bias occurs when service outcomes are unfavorable. However, we have not determined whether positive biases toward employees with certain accents exhibit similar trends. In addition to negative accent biases, research has found that some accents, such as the standard British accent, are viewed positively (Cargile, 2000; Stewart et al., 1985). Positive biases are less well understood because of the widespread belief that they are less detrimental to society (Crandall & Eshleman, 2003). However, understanding accents receiving a positive bias is important for managers because a positive bias could enhance customer outcomes and, accordingly, could influence a variety of call service decisions.

We anticipate that customers in the U.S. will have a positive bias toward employees with British accents because a British accent has conventionally been associated with prestige and professionalism (Stewart et al., 1985). That is, customers are likely to evaluate British-accented employees more favorably and experience more positive feelings while interacting with them compared with interactions with American- or Indian-accented employees.

Unlike negative accent biases, positive biases are unlikely to be suppressed. Suppression requires mental effort, so people only suppress biases when they believe it is important to do so. Because positive biases seldom feel disruptive, we expect that they are suppressed only when the contextual information clearly contravenes such positive accent-based stereotypes. For example, an employee’s British accent should evoke favorable perceptions, and this positive prejudice is likely to be experienced and expressed in its genuine form. However, if the service delivered by this British-accented employee does not provide favorable outcomes for customers, the customers will suppress their positive reactions. Hence, when a service outcome is unfavorable, accents receiving a negative bias will be rated lower than standard accents and accents receiving a positive bias will lose their advantage over standard accents. Formally,
H4. Customers will rate employees with an accent receiving a positive bias higher than employees with a standard accent only when a service outcome is favorable for customers.

5.2. Attritions

Consistent with our earlier findings, we expect that the result of the service outcome will influence customer attributions. Customer attributions of service outcomes can be dispositional, that is, assigned to the inherent traits of the actor, or situational, that is, assigned to external conditions (Cowley, 2005). Research on stereotype maintenance shows that people attribute stereotype-consistent behaviors to internal, dispositional causes but stereotype-inconsistent behaviors to external, situational causes (e.g., Crocker, Hannah, & Weber, 1983). We predict that people will err toward dispositional attributions for unfavorable outcomes when accents are negatively perceived and for favorable outcomes when accents are positively perceived. Formally,

H5. When the outcome of the service outcome matches the accent-based bias (e.g., Indian-accented employee-negative outcome, British-accented employee-positive outcome), customers are more likely to attribute the outcome to the employee rather than to external causes.

5.3. Information about typical outcomes

According to Szymanski and Henard (2001), one of the most important antecedents of customer satisfaction is the degree to which customers feel they have been treated equitably, which they estimate by making comparisons with other customers. However, customers often cannot observe other customers’ service interactions, particularly in call service encounters. When customers cannot make reliable comparisons, we expect that they will depend more on stereotypical information to make service evaluations. Therefore, providing customers with information about the treatment of other customers should reduce their reliance on stereotypical information to make service evaluations. Therefore, providing customers with information about the treatment of other customers should reduce their reliance on stereotypical information. Previous research on stereotyping has shown that giving people more information about a situation can reduce the effects of stereotyping (Kunda & Thagard, 1996; Locksley, Borgida, Brekke, & Hepburn, 1980). Thus, we propose that providing information about the frequency of an unfavorable vs. a favorable service outcome will decrease the customer propensity to use accent stereotypes. For example, if customers know that outcomes are usually unfavorable, customers will be less likely to rely on stereotypical information about the employee. Formally,

H6. Accent biases are less likely to impact service evaluations if customers have information about the frequency of a favorable vs. unfavorable service outcome.

5.4. Design, participants, and procedure

As before, we used an Indian accent for the negative bias and a standard American accent as the standard. Consistent with the findings of Stewart et al. (1985), we used the standard British accent to represent the positive bias. Study 3 used a 3 (employee accent: Indian, British, and standard American) × 2 (service outcome: favorable vs. unfavorable) × 2 (information about typical service outcome: available vs. unavailable) between-subjects design. Two hundred ten students at a large U.S. Midwestern university, who participated in the experiment in exchange for extra credit, were randomly assigned to one of the twelve experimental conditions. Consistent with Study 2, we adopted a mood-neutralizing procedure to ensure that participants’ pre-encounter moods were comparable.

The service scenario in this study is a customer calling to request the cancellation of overdraft charges. We manipulated the information availability by providing or withholding information pertaining to the frequency of favorable vs. unfavorable outcomes. Specifically, when the information about the typical service outcome is available, customers are told that it is common for banks to charge an overdraft fee as a condition of their contract. The average of an overdraft charge is $33. Only 10% of the overdraft fees are ever reimbursed based on customer request. In the conditions where information about the typical service outcome is unavailable, participants read the same introduction about overdraft charges, but the information that “Only 10% of the overdraft fees are ever reimbursed” is not provided to the customer.

5.5. Dependent measures

For Study 3, we used two dependent variables, customer satisfaction and post-service attribution. Customer satisfaction was measured using the 4-item, 7-point Likert scale from Hennig-Thurau et al. (2006). To measure post-service attribution, respondents were asked to describe their experience with the service encounter using an open-ended response. Then, two independent coders judged whether the respondents attributed the service outcome to the employee’s disposition

![Fig. 4. Customer satisfaction as a function of employees’ accent, service outcomes, and information context in Study 3. Notes: Error bars indicate 95% confidence intervals.](image-url)
(e.g., employee competence or care for the customer) or external causes (e.g., organizational policy or luck).

5.6. Manipulation check

In this experiment, we asked participants to guess the employee’s accent. Most participants identified the accent correctly (86% for the Indian accent, 96% for the American accent, 91% for the British accent), which provided further confirmation that our manipulation worked. The mood neutralizing procedure was also successful because customers’ pre-encounter moods did not differ significantly across experimental conditions \((F(1, 198) = .12, \text{NS})\).

5.7. Results

5.7.1. Customer satisfaction

We constructed the customer satisfaction index by averaging the four relevant items (coefficient \(\alpha = .96\)). The means and deviations appear in Table 4. The results of a \(3 \times 2 \times 2\) ANOVA of customer satisfaction (Fig. 4) revealed a main effect of service outcomes \((F(1, 198) = 301.98, p < .01)\) and the predicted three-way interaction of accent, service outcome, and information \((F(2, 198) = 2.96, p = .05)\). Accent stereotypical effects should have decreased when customers were informed about the industrial norms. Thus, in support of H6, the follow-up tests for each information context revealed that when no other objective information was available, the results were consistent with those from Study 1. The main effect of the service outcome \((F(2,85) = 229.48, p < .01)\) was qualified by a significant two-way interaction between accent and service outcomes \((F(2,85) = 4.37, p = .02)\). Planned contrasts revealed that when the service outcome was favorable, a British accent led to significantly higher customer satisfaction than did the USA accent \((M_b = 5.70, M_A = 4.76, F(1, 85) = 6.16, p = .02)\). Customer satisfaction in the Indian accent condition did not differ significantly from that in the standard American accent condition \((M_I = 5.20, M_A = 4.76, F(1, 85) = 1.67, F = 1, \text{NS})\). However, when the service outcome was unfavorable, customer satisfaction ratings were not notably different in the British and standard American accent conditions \((M_b = 2.21, M_A = 2.52, F = 1, \text{NS})\), whereas the Indian accent condition featured significantly lower customer satisfaction ratings \((M_I = 1.67, M_A = 2.52; F(1, 85) = 5.94, p = .02)\). Thus, the results support H6 and replicate H7. When customers had other objective information to evaluate their service outcomes (knowledge about typical service outcome), only the main effect of the service outcome was significant \((F(1, 113) = 119.89, p < .01; \text{all other Fs} = \text{NS})\). That is, customer satisfaction ratings were not significant when customers had more objective information. Thus, H6 is supported.

5.7.2. Service attributions

We again analyzed customers’ attributions of the service outcome, but instead of asking participants directly about their employee attributions, we requested that they write open-ended responses to an oblique question: “If you were to describe your experience of this service encounter to your friends, what would you say?” Two judges, unaware of the research purpose, independently coded the responses and resolved any differences through discussion, although they agreed on 96% of the responses. The responses could be assigned to one of two categories: the employee’s disposition (e.g., competence, warmth, caring for customers) and external factors (e.g., bank policy, customer luck). With a binary logistic regression, we regressed attribution (external = 0, dispositional = 1) on employees’ accents (American = 0, Indian = 1, British = 2), service outcomes (favorable = 0, unfavorable = 1), the information context (no other information provided = 0, objective outcome information provided = 1), and their interactions. The test of the model was significant \((\chi^2 = 39.18, p < .01)\), although the three-way interaction was not \((\beta = -.24, \text{NS})\). Both the accent \(\times\) outcome interaction \((\beta = -1.98, p = .02)\) and the service outcome \((\beta = -2.29, p = .04)\) were significant. Following Wood’s (2010) example, we analyzed the model using indicator contrasts of the variables. When the service outcome was unfavorable, the probability that customers would attribute the outcome to employee disposition increased in the Indian accent condition compared with the American accent condition \((\beta = -1.54, p < .01)\), but this effect was not evident when the service outcome was favorable. Instead, when the service outcome was favorable, the probability that customers would attribute the outcome to employee disposition increased in the British accent condition compared with the American accent condition \((\beta = -.99, p = .05)\), and this relationship also disappeared when the service outcome was unfavorable. These findings support H6.

5.8. Discussion

Using a different service context (i.e., request for overdraft fee cancellation), Study 3 provides further support for the justification and suppression model. As before, when a service outcome is unfavorable, accents receiving a negative bias are rated lower than standard accents. However, when an outcome is favorable, accents receiving a positive bias lead to higher ratings than standard accents. Additionally, both positive and negative accent stereotyping effects decrease when customers have other diagnostic cues – in this case, knowledge about the frequency of an unfavorable vs. favorable service outcome – to influence their judgments.

6. Conclusions and implications

Given the prevalence of employees with nonnative accents providing customer service by phone, the paucity of research examining accent biases in call service encounters is surprising. Even without sufficient research, many companies are adopting costly measures to reduce negative customer bias toward nonstandard accents, such as relocating call centers and implementing speech training programs. Using self-reported measures and open-ended questions in both lab and field studies, our research addresses this gap by (1) applying the JSM to call service interactions to explain when accent biases impact customer evaluations, (2) exploring the underlying mental mechanisms of suppressing and justifying accent biases, and (3) providing evidence that not all accent biases are negative.

First, to our knowledge, our study is the first to apply the JSM framework to a service encounter. To date, the consumer research into customer biases regarding employees and psychological research on the suppression and justification of prejudice have existed more or less independently. Most customer behavior research has focused on how consumers use stereotypes to make judgments about products, service providers, or fellow consumers. For example, customers stereotype employees on the basis of their age (Kang & Hillery, 1998), gender (Matta & Folkes, 2005), race (Jones, Moore, Stanaland, & Wyatt, 1998), and appearance (Kang & Herr, 2006). However, research has not examined whether customers also suppress or justify their prejudices toward employees, particularly in call service encounters. Our research shows that the justification and suppression model applies even in a relatively impersonal phone service context. This important finding means that the compulsion to justify or suppress prejudice is not limited to face-to-face service encounters in which social norms or political correctness discourage negative prejudice.

Second, our research shows that the suppression of negative biases requires both cognitive and affective effort. The role of affect in suppression is particularly interesting because it means that when customers are upset, they are less able to suppress their negative biases. According to Munichor and Rafaeli (2007, 517), when providing call service, “organizations cannot avoid making customers wait on hold because the required costs would be prohibitive.” Munichor and Rafaeli (2007) find
that messages apologizing to customers for the hold time results in the most negative customer reactions, while messages providing customers with information about their location in the queue leads to the most positive customer reactions. Hence, before even speaking to a call service employees, customers are often already annoyed or agitated. We argue that when customers begin a service call already frustrated, they will be less able to suppress negative biases. Therefore, mechanisms for reducing negative accent biases should begin prior to the customer–employee interaction.

Finally, although previous research has been unable to provide convincing evidence that people may be positively biased toward outgroups (Mullen, Brown, & Smith, 1992), our results indicate that customers can be positively predisposed toward certain accents. One possible reason why previous research has not found consistent evidence of positive biases toward outgroups is that negative and positive biases manifest under different circumstances. In our study, positive accent biases only influence customer evaluations when outcomes are favorable; when outcomes are unfavorable, the benefits of positive biases disappear. Thus, although people suppress their negative prejudices unless evidence justifies them, people do not suppress positive prejudices unless there is evidence that contradicts them.

### 6.1. Managerial implications

By exploring the triggers of accent prejudice in call service settings, we derive three main actionable recommendations for reducing prejudice against employees with nonstandard accents. First, customers in the U.S. suppress their negative prejudice, but not their positive prejudice, when service outcomes are favorable. Therefore, services that are likely to result in a favorable outcome for customers are more appropriate assignments for employees with nonstandard accents than services that are likely to lead to unfavorable outcomes. It might be tempting to outsource unpleasant and time-consuming tasks to international call centers to reduce costs, but the prevalence of employees with nonstandard accents could cause customers to become more upset than they would have been if they had dealt with standard-accented employees. Even using employees with accents receiving a positive bias to relay bad news will be ineffective because customers suppress their positive bias when they confront unfavorable outcomes. Hence, activating new accounts or providing good news represents better tasks for nonstandard-accented employees than relaying bad news or providing service where the probability of an unfavorable outcome is high.

Second, call service managers should work diligently to provide high-quality, favorable results for customers whenever they can. Our results show that customers in the U.S. suppress their negative accent biases and maintain their positive ones as long as they feel a service outcome is favorable. It is noteworthy that many employees residing in the U.S. also have nonstandard accents, so even domestic businesses providing service by phone will benefit from ensuring that customers receive favorable outcomes.

Finally, providing customers with transparent and accurate information reduces their use of prejudicial information in making service outcome attributions. We therefore extend the findings of Roggeveen et al. (2007): providing information about firm reputation reduces prejudice against call service employees. When using nonstandard-accented employees to provide customer service by phone, managers should provide customers with information about industry norms, such as wait times, conditions in which service contracts can be terminated, and customer rights. This information must be based firmly in facts; otherwise, customers will disregard it. Customers base their expectations on their own or others' previous experiences, so they likely can recognize when they receive inaccurate or misleading information. For example, some firms tell customers that they never waive certain fees, even if waiving or reducing fees is commonplace in the industry. Instead, managers should inform customers about when fees can be waived because even if the customers do not meet those conditions, they are less likely to make damaging attributions when they have such information.

### 6.2. Limitations and further research

While this research was limited to consumers in the U.S., businesses hire nonstandard-accented employees to provide call service in many other countries, as well. As such, we recommend that future research explore accent biases in call service encounters in other countries and cultures. Additionally, we used the JSM to explain accent bias in the call service setting. However, it would be interesting to apply the JSM to other types of customer biases and in other customer–employee settings. Further, we only investigated one mechanism, information availability, for decreasing customer bias when the service outcome was unfavorable. Research extensions should determine whether there are other mechanisms that also decrease customer bias. Finally, we used the Indian accent in our lab studies and the Indian accent was the most common non-Western accent reported in the field test. Although our findings apply to other accents, it should be noted that the biases toward them may be weaker or stronger than the Indian accent. It is possible that a negative bias could be so weak that it is not managerially relevant even when the bias is not suppressed. Hence, even in light of our findings, knowing the strength of accent bias remains important.

### Appendix A

**Items for measurement scales.**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer–employee rapport (Adopted from Hennig-Thurau et al., 2006)</td>
<td>I enjoyed interacting with this employee. This employee created a feeling of “warmth” in our interaction. This employee related well to me. I was comfortable in interacting with this employee.</td>
</tr>
<tr>
<td>Customer satisfaction (Adopted from Hennig-Thurau et al., 2006)</td>
<td>I was delighted by this service experience. I was satisfied with this specific service experience. I really liked this service experience.</td>
</tr>
<tr>
<td>Changes in customer’s mood (Adapted from Tsai &amp; Huang, 2002 and Hennig-Thurau et al., 2006)</td>
<td>Warmhearted Pleased Cheerful Upset Angry Annoyed</td>
</tr>
<tr>
<td>Evaluation of employee performance (Adapted from Cornin, Brady, and Hult 2000)</td>
<td>This service employee was willing and able to provide service in a timely manner. This service employee was competent. This service employee was approachable and easy to contact. This service employee was easy to understand. This service employee was trustworthy. This service employee made the effort to understand my needs. This service employee was friendly. Please let us know your explanation for the level of service that the service provider delivered to his or her customers. If you were to describe your experience of this service encounter to your friends, what would you say?</td>
</tr>
</tbody>
</table>

---

*We would like to thank Reviewer 1 for this suggestion.*
References


Gremler, D. D., & Gwinner, K. P. (2000). Customer...
Marketing function and form: How functionalist and experiential architectures affect corporate brand personality

Ursula Raffelt a, Bernd Schmitt b,c, *, Anton Meyer a

a Munich School of Management, Ludwigstr. 28 RG III, 80539 Munich, Germany
b Columbia Business School, 3022 Broadway, New York, NY 10027, United States
c Institute on Asian Consumer Insight (ACI) at Nanyang Technological University, Singapore

A R T I C L E   I N F O

Article history:
First received in 8, March 2012 and was under review for 5 months
Available online 16 May 2013
Area Editor: Amitava Chattopadhyay

Keywords:
Corporate architecture
Design
Corporate image
Brand personality
Experience

A B S T R A C T

How are the designs of corporate buildings used to create meaning and project a corporate image and personality? We distinguish functionalist architecture (“form follows function”), which focuses on the primary, utilitarian function of a building, from experiential architecture (“from function to form”), which uses the form of a building to communicate symbolically about the organization. A large-scale quantitative study including 150 buildings shows that four architectural design types (“solid,” “balanced,” “expressive,” and “disruptive” designs, emerging from a mix of functionalist and experiential architectures, lead to distinct corporate brand personalities (e.g., competence for functionalist architecture and excitement for experiential architecture). We validate these findings in a qualitative study and discuss how this research contributes toward the development of a consumer-oriented design theory.

© 2013 Published by Elsevier B.V.

1. Introduction

Corporate buildings are omnipresent in our daily lives. We encounter these buildings by sight, by visiting them, and through the media. Corporate buildings are the physical manifestations of the organizations within them and shape and define our landscapes and the skylines of our cities. Most importantly, in today’s consumer society, many organizations use corporate buildings to project a corporate image and personality to their current and potential customers.

Prior research has largely neglected the issue of the relation between corporate architecture and brand personality. Marketing scholars have established that visual expression is a form of communication and have shown that design elements contribute to the formation of brand beliefs and influence brand strength. This research focused largely on micro-level issues of design and meaning, examining logos (Henderson & Cote, 1998, p. 23), typefaces (Henderson, Giese, & Cote, 2004, p. 64), packaging (DeBono, Leavitt, & Backus, 2003; Leder, Carbon, & Kreuzbauer, 2007; Orth & Malkewitz, 2008; Underwood, 2003; Underwood & Klein, 2002) and product design (Brunel, 2006).

Architectural design, however, is far more complex than any of the prior design stimuli investigated. Architectural language comprises a continuum, ranging from very concrete, construction-related, technical terms to abstract, impression-related, stylistic terms. It is therefore important to examine how consumers view architectural stimuli, such as corporate buildings and company headquarters, and how companies may use their buildings to create meaning and project a corporate image and personality. Moreover, insight into the complex designs in architecture may add to the research findings emerging from simpler designs found in logos, typefaces and products.

In this article, we distinguish two general design types of corporate architecture that we expect to create corporate brand impressions and personalities. We present a large-scale quantitative study that relates architectural design dimensions (e.g., degree of elaborateness or harmony) and design types (e.g., functionalist and experiential) to corporate brand personality. We subsequently validate and extend our findings through a qualitative study. In both studies, we focus on the external shape rather than the interiors of corporate buildings because consumers define an image of a company based on the exterior appearance.

2. Two key architectural design types and their variations

The rise of the modern organization following the industrial revolution of the nineteenth century was accompanied by architectural styles that reflected the spirit of the times. Originally, corporate architecture followed the principle “form follows function” that was first formulated by American architect Louis Sullivan in 1896. Sullivan’s principle, which became closely associated with the functionalist
movement in architecture, states that the shape—or “form”—of a building should be based solely on its intended function or purpose. The architects of the influential Bauhaus school, such as Walter Gropius and Mies van der Rohe, popularized the functionalist view of architecture, which influenced architectural design throughout most of the 20th century, and into the 21st century, through the International Style (Le Corbusier and, later, Philip Johnson) and the works of I.M. Pei and Richard Meier.

As the economy shifted from production to consumption and toward an “experience economy” (Pine & Gilmore, 1999) and as marketing shifted from a focus on features and benefits toward “experiential marketing” (Holbrook & Hirschman, 1982; Schmitt, 1999), the established functionalist language of corporate architecture gave way to a new aesthetic, emphasizing expression, symbolism, the plurality of form, and experience (Venturi, Brown, & Izenour, 1977), which we will henceforth call “experiential architecture.” Experiential architecture constitutes a shift “from function to form” (Klingmann, 2007), with corporate architecture now emphasizing the brand instead of the functionality of the building (Bahamón, Cañizares, & Corcuera, 2009). Corporate buildings are viewed explicitly as a form of communication (Hattenhauer, 1984). While the functionalist style of architecture is still widespread, experiential architecture is increasingly being used in the design of new corporate buildings today. Experiential architecture is associated with the postmodern and deconstructivist movements and with architects such as Frank Gehry, Rem Koolhaas, Daniel Libeskind, and Zaha Hadid.

Well-known functionalist and experiential buildings are shown in Fig. 1 a. Some of these buildings are considered classics; others are fairly recent buildings. Examples also refer to different movements, e.g., the early industrial orthogonal look of the Bauhaus; the rounder, more natural looking buildings of the later phases of functionalist architecture; and the experiential architecture of the so-called “blobitecture” and deconstructivist movements. Fig. 1 b contrasts functionalist and experiential buildings of three organizations—BMW, a German car manufacturer; CCTV, mainland China’s television broadcaster; and Columbia Business School, a business school in New York. BMW Welt, a multi-purpose facility built in Munich by the renowned architectural firm CoopHimmel[1]au in 2007, uses visual analogies of a tornado followed by a cover of clouds to elicit consumer perceptions of dynamism and challenge—the core brand values of BMW (Feireiss & Kwinter, 2007). In comparison, the BMW headquarters, completed in 1972 in the midst of the modernist movement, is a functionalist office building, visually representing components of automotive engines. Similarly, star architect Rem Koolhaas’ new building for CCTV, completed in Beijing in 2008, is supposed to express the value of “collective inhabitation” (Zalewski, 2005), and the newly planned building of Columbia Business School in New York City is described by the architects as “an open and inter-disciplinary model [that] replaces the top-down logic of industrial-age knowledge transfer” (http://www.unstudio.com/research/iop/beyond-the-classroom-research-on-knowledge-spaces). In contrast, CCTV’s earlier headquarters and Columbia Business School’s 1961 building are purely functionalist designs.

We view functionalist and experiential architectures as “architectural ideal types,” emerging from multiple building design dimensions, including, for example, a certain degree of elaborateness, harmony, transparency, and colorfulness. Depending on the exact values on each dimension, an architectural design may be perceived as more or less typical of the functionalist or experiential categories. As a result, in our empirical studies, more than two design types may emerge. For example, based on the specific values of the dimensions, there may be a stark, solid design type that is largely functionalist and a softer version that uses functionalist elements in a more relaxed and balanced way. Similarly, there may be an exaggerated, deconstructive and disruptive experiential design type and a softer, still expressive, yet less aggressive experiential design type.

3. Research framework

Which brand images and personalities might be associated with which architectural design dimensions and design types? In a seminal essay on the semiotics of architecture, the Italian semiotician Umberto Eco distinguished between what he calls the “denotation” of a building (its primary utilitarian function) and the “connotation” of a building (its symbolic meaning) (Eco, 1997). Functionalist architecture, with its simplified and proportional forms, horizontal and vertical lines, and stark, unornamented, rational, and industrial look (Le Corbusier, 1968; Wolfe, 1981), focuses primarily on the denotative function. In contrast, experiential architecture, with its eclectic forms, multiple references, and complex, ornamental, and playful design elements, which are inspired by postmodern ideas (Jencks, 1987), emphasizes the connotative function.

Indeed, functionalist architectural designs gained prominence during early market capitalism, when the modern corporation began to use rational rules with homogenous analytical procedures and implemented a model of production, utility, and efficiency (Weber, 1922/1978), resulting in a “disenchantment of the world” (Weber, 1922/1978). Functionalism thus stood for the capitalist philosophy and projected an overall image of rationality and utilitarianism in its architecture (using denotative elements such as vertical lines and a rational, industrial look). With the emergence of a postmodern society, society and the organization became “(re)enchanted” (Firat & Venkatesh, 1995; Jenkins, 2000). In addition to rational elements, organizations began to stress hedonic, emotional, creative, and innovative elements (Balmer & Greyser, 2003; Gobe, 2001), and the emerging experiential designs (using connotative elements and an ornamental and playful look) reflected this change in organizations.

As a result, we expect that people form different brand personality impressions about organizations based on the organizations’ corporate architecture. Brand personality—defined as the “set of human characteristics associated with the brand” (Aaker, 1997, 347)—captures trait-like associations and inferences about commercial symbols. Perceptions of brand personality can be formed by any direct and indirect contact that a consumer has with a brand (Aaker, 1997; Johar, Sengupta, & Aaker, 2005). Research has shown that brand personality impressions result not only from exposure to the brand name and advertising but also can be formed on the basis of product design (Govers & Schoormans, 2005), typeface design (Henderson et al., 2004), package design (Orth & Malkewitz, 2008), and retail design (D’Astous & Lévesque, 2003; Martineau, 1958). Upon viewing photos of the interior and exterior of homeowners’ dwellings, respondents were able to infer the personality of the home owners (Sadalla, Vershure, & Burroughs, 1987). Similarly, we expect that people infer the personality of an organization when shown an image of its corporate building.

While corporate architecture may affect all dimensions of brand personality, we expect two personality dimensions—competence and excitement—to be particularly relevant and formative in the consumer representations and images of the architectural dimensions and designs. The competence dimension, which is strongly associated with traits such as reliable, responsible, dependable and efficient (Aaker, 1997, 351), is conceptually closely related to the ideas of predictability, efficiency and rationality, which following Weber (1922/1978) are essential elements of early capitalist society and inherent values of functionalist architecture. In contrast, the excitement dimension, which is associated with traits such as excitement, imagination, spiritedness and trendiness (Aaker, 1997, 351), appears closely related to postmodernism and is expressed in experiential architecture. Thus, we expect design dimensions and design types to affect corporate personality. In particular, we predict that empirically emerging design dimensions that are characteristic of functionalism (for example, proportional horizontal/vertical lines and lack of elaborateness) and empirically emerging functionalist design types should
evoke an impression of competence. In contrast, design dimensions that are characteristic of experiential architecture (such as elaborate designs) and experiential design types should evoke an impression of excitement.

To summarize, the purpose and contribution of this article is to address two interrelated empirical questions. First, which architectural design dimensions are empirically related to which personality dimensions? Second, which design types are related to which personality dimensions? Answers to these questions are not only of theoretical interest; they also provide guidance for architects and marketers. Architects may become aware of the specific marketing impact of their designs on consumers, and marketers may ask architects for certain designs if they intend to project a certain corporate personality.

4. Study 1: relating architectural design dimensions and design types to corporate personality

Study 1 examines how architectural dimensions and styles project corporate images and personalities through design. The study was
conducted as a quantitative, mixed-sample online survey with architectural experts and consumers and included 150 buildings from around the world.

Following prior marketing research on visual design that has related design attributes and dimensions of typefaces and packaging to consumer impressions (Henderson et al., 2004; Orth & Malkewitz, 2008), we measured specific design attributes (e.g., the color or form of a building) and extracted broader architectural design dimensions (elaborateness or harmony) that are part of a building’s architecture. We then created architectural design types (both functionalist and experiential) and correlated the design dimensions and types with different corporate personality impressions. Because we used a large sample set of buildings, this study enabled us not only to examine design types at a general and broad level as functionalist and experiential designs but also to test whether there are other design types and whether there are, perhaps, subtypes of functionalist and experiential designs.

4.1. Method

A total of 652 architectural experts (i.e., architects and architecture students) evaluated 150 buildings in terms of their design attributes, and 566 ordinary consumers (i.e., non-architecture students) evaluated the same 150 buildings in terms of their corporate personality impressions. The buildings were presented on a computer screen next to the items. Each building was presented individually on the left side of the screen; the respondents scrolled down the scale items on the right. This procedure ensured that the building was in full view throughout the rating process. The experts rated an average of two to three randomly assigned stimuli. The task took an average of 18 min. In total, the 652 experts provided 1934 design ratings. Each building received an average of 13 participant evaluations. Following the methods described by Holbrook and Batra (1987), we assessed inter-rater reliability by computing coefficient alphas for each of the 64 design attributes. We regarded the expert ratings as the “items” and computed coefficient alphas across the ratings for each attribute. Inter-rater reliabilities averaged .78 with a median of .82, which provides justification for further analyses at the stimulus level.

The non-expert consumers rated an average of three to four randomly assigned stimuli. In total, they provided a total of 1917 brand personality evaluations. Again, each building stimulus was evaluated by approximately 13 participants, with inter-rater reliabilities averaging .65 (median of .72), which was considered acceptable. As an incentive, the respondents participated in a lottery to win prizes.

4.1.1. Architectural stimuli

To include a broad range of buildings as stimuli, we reviewed architectural publications (Ielonart, 2009; Messedat, 2005), conducted extensive Internet research and consulted professional architects. An initial pool of 100 buildings was generated. In a pretest, we asked eight architects and architectural critics to rate the stimuli (25 buildings per architect) on 64 building design attributes (see below). Because not all attributes appeared to be well represented by the stimuli and because the variance among the buildings was not sufficiently large, we added another 50 buildings to the initial set. In the actual study, these 150 corporate buildings were presented as two-dimensional, high-resolution images. Company names were not mentioned to avoid confounding architectural evaluations with name-induced impressions. The images were standardized regarding illumination, image details, image section, and size (a width of 340 pixels, a resolution of 72 dpi).

4.1.2. Measurement of primary and secondary design attributes

Architects and architectural critics have described building designs at various levels of abstraction. For the purpose of mapping meaningful architectural design descriptions to personality associations, it appears most appropriate to describe architectural designs at intermediate levels to avoid the extremes of excessively detailed descriptions (such as the pitch of a roof) as well as highly abstract descriptions (such as a particular era), both of which may be useful to architects but not to marketing researchers. Intermediate-level descriptions in the architectural literature include impression-like bipolar items of design attributes that are linked to physical attributes of architecture; these so-called primary design attributes include color (e.g., “warm vs. cold”), material (e.g., “natural vs. artificial”), form (e.g., “technological vs. organic”), and facade (e.g., “simple vs. complex”). In addition, intermediate-level descriptions include impression-like semantic descriptions of the architecture as a whole; these so-called secondary design attributes include “calm vs. lively,” “exclusive vs. ordinary,” and “urban vs. rustic.” A total of 159 intermediate-level items were initially selected from architectural research (Canter, 1969; Devlin & Nasar, 1989; Ellis, 1993; Gifford, Hine, Muller-Clemm, Reynolds, & Shaw, 2000; Hershberger & Cass, 1992; Joedicke, Dirlewanger, Geisler, & Magnago-Lampugnani, 1977; Karlsson, Aronsson, & Svensson, 2003; Vielhauer Kasmar, 1992). Through several rounds of pretesting with architects and architectural students, the number of items was reduced. The final scale consisted of 64 items.

4.1.3. Measurement of corporate brand personality

Brand personality was measured on a 44-item trait scale. The scale was broadly based on prior brand personality research; approximately half of the items were drawn from Aaker (1997). Several items were added from D’Astous and Lévesque’s (2003) store personality scale. Some items were reworded and adapted, and some items were added to fit the architectural context.

4.1.4. Additional measures

Architectural knowledge and expertise was assessed based on a self-report measure by Flynn and Goldsmith (1999). The mean competence was significantly higher for experts (M = 5.18) than non-experts (M = 2.55), p < .001. In addition, experts and non-expert consumers rated each building on two separate 7-point scales in terms of its plausibility as a corporate building and its overall aesthetic appeal. The responses of participants in the architecture experts sample that indicated that they were not expert architects were excluded, as were the responses of participants in the non-experts sample that indicated that they were experts in design fields. Furthermore, the responses of participants who indicated high recognition values for a building (an evaluation of 6 or 7 on a 7-point recognition scale and/or a correct identification in an open-ended question) were excluded because building recognition could bias the attribution of brand personality traits.

4.2. Data analysis

The data analysis was based on procedures previously established in visual design-related research (Henderson & Cote, 1998; Henderson, Cote, Leong, & Schmitt, 2003; Henderson et al., 2004; Orth & Malkewitz, 2008). All analyses were conducted at the stimulus level. To obtain a score for each stimulus, we computed the mean of each building’s individual ratings. All remaining analyses were conducted using such mean scores. This procedure allows both datasets to be integrated. Hence, the sample size of each analysis is the number of buildings rated by experts and consumers (n = 149). One stimulus was excluded because it had much higher recognition ratings than the other stimuli.

4.3. Results

4.3.1. Creating design dimensions and brand personality dimensions

Conducting two separate factor analyses, we first created broad design dimensions and brand personality dimensions. Using the “elbow” criterion, a five-factor solution emerged for the design dimensions, explaining 62% of the variance. The first factor, interpreted as “elaborateness” based on its highest factor loadings, encompasses
mostly secondary design attributes that are perceived as personal, unique, and imaginative. Elaborateness is not considered banal, monotonous or rational but instead is considered exclusive, striking, expressive, playful, and progressive. On the primary design attribute level, elaborateness is characterized by free-flowing forms and a nonfunctional façade. The second factor, interpreted as “harmony,” is considered harmonious, comforting, and coherent. Harmony is also considered clear and elegant. At the primary level, a proportional building form is the highest-loading variable on this factor. The third factor, interpreted as “natural feel,” is characterized mostly by two primary design-attribute types: materials (natural, absorbing, rough, unrefined) and colors (dull, warm). Regarding secondary design attributes, this factor received high ratings for rustic and cozy attributes. The fourth factor, labeled “transparency,” is considered open in its façade and transparent in its use of materials. This factor is not weightless and graceful. Finally, the fifth factor was interpreted as “colorfulness” because it received high loadings from several color-related primary design attributes.

The factor analysis conducted on the corporate personality traits revealed a four-factor solution, explaining 83% of the variance. The four dimensions that emerged from our research—excitement, competence, sophistication, and sincerity—corresponded closely to Aaker’s (1997) brand personality dimensions; namely, excitement, competence, sophistication and sincerity. In particular, the items with high loadings on the first factor, which was interpreted as “excitement,” referred to arousal- and change-related items (such as exciting, forward-looking, revolutionary, unique, imaginative, lively). The second factor, which was interpreted as “competence,” included items with high loadings on reputation (reputable, competent) and reliability (technical, responsible, reliable). The third factor, which was interpreted as “colorfulness,” included items referring to repetition and appeal from an aesthetic and high-end point of view (glamorous, elegant, charming, good-looking, upper-class). Finally, the fourth factor, which was interpreted as “naturalness” (in both a concrete and more abstract sense of honesty and sincerity), included four items with high loadings (down-to-earth, natural, honest, and real).

4.3.2. Relating design dimensions and personality dimensions

A canonical correlation analysis was conducted to examine whether, overall, the design and brand personality dimensions are significantly related. The full model across all functions was statistically significant (Wilks’ $\lambda = .051$, $F(20, 465.28) = 33.65$, $p < .001$). Multiple linear regression models were applied to assess each design dimension’s contribution to the consumers’ impressions, using the design factors as predictors. All four regression models were significant at $p < .001$. The explained variance of the brand personality dimensions was large, with an adjusted $R^2$ ranging between .421 and .583, except for naturalness ($R^2 = .175$). The results are shown in Table 1.

Considering the regression coefficients, excitement is most strongly predicted by elaborateness in design, followed by lack of natural feel and harmony as well as transparency. Competence is predicted by harmony, lack of elaborateness, and lack of colorfulness. Stylishness is predicted by lack of natural feel and elaborateness. Finally, naturalness is predicted by all design dimensions but is most strongly predicted by a natural feel in the design.

4.3.3. Creating architectural design types

Next, we identified architectural design types based on the similarities of primary and secondary design attributes emerging from experts’ ratings to see whether functionalist and experiential designs emerge as major architectural design types and whether there are other design types or subtypes of functionalist and experiential designs.

A two-step cluster analysis, which considers holistic perceptions of visual designs, was used to form homogeneous design types. The optimal number of clusters was determined by Akaike’s information criterion. Based on a log-likelihood distance measure, a four-cluster solution of prototypical architectural designs emerged. The cases were relatively evenly distributed among the clusters.

To facilitate the interpretation of the design clusters, the design factors (or dimensions) were used to describe and interpret the previously identified design clusters. A MANOVA, with design factor scores as dependent variables and the design clusters as the independent variable, indicated that, overall, the design dimensions differentiated the design clusters (Wilks’ $\lambda = .085$, $p < .001$). Subsequent ANOVAs (all $p$ values $< .01$) followed by t-tests provided information on the factor scores that differed significantly from the mean factor score across all clusters. The strongest design differentiators were elaborateness and harmony, while the weakest differentiating factor was colorfulness. Table 2a shows the four architectural design types that emerged from the cluster analysis and the corresponding design dimensions. Table 2b shows the relation between design types and brand personality dimensions.

<table>
<thead>
<tr>
<th>Design dimensions</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elaborateness</td>
<td>-.06</td>
<td>.25</td>
<td>-.04</td>
<td>.25</td>
</tr>
<tr>
<td>Harmony</td>
<td>-.81</td>
<td>.00</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>Natural feel</td>
<td>-.05</td>
<td>.52</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Transpency</td>
<td>-.04</td>
<td>.56</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>Colorfulness</td>
<td>.17</td>
<td>.32</td>
<td>-.33</td>
<td>-.33</td>
</tr>
</tbody>
</table>

Notes. Design dimensions and brand personality dimensions are $z$-standardized ($M = 0$, $SD = 1$). T-tests refer to differences from the sample means (bold indicates $p < .05$; * indicates $p < .1$; *** indicates $p < .01$). In addition to $t$-tests, within each brand personality dimension, different superscripts identify pairs of designs that score significantly different ($p < .05$) on this factor, on the basis of a Tamhane post hoc. Positive numbers indicate positive values on each dimension, e.g., high elaborateness or high excitement.
brand personality dimensions, which will be discussed later. Table 2c provides typical examples of buildings for each cluster.

Cluster 1 is distinct from all other clusters based on its below-average harmony scores. All the other factors are not significantly different from the z-standardized sample mean of 0. As we know based on the factor loadings of the primary and secondary design elements discussed earlier, this lack of harmony means that the buildings in this category are perceived as dissonant, confusing, clumsy, improvising, and faddish. The buildings are associated with negative evaluations, such as intimidating and threatening. The building shapes are not seen as proportional. This cluster, which included 45 buildings, may be interpreted as representing the architectural design type of “disruptive design” and includes several examples from the deconstructivist movement (see Table 2c).

Cluster 2 is differentiated on the elaborateness dimension, which is above average, and the natural feel dimension, which is below average. These buildings evoke a personal, unique, and imaginative impression, and they are considered intriguing, exclusive, and lively. The building form is free-flowing and nontechnical, and the façades are nonfunctional and seen as three-dimensional. The harmony values are slightly below average. In terms of the primary design elements within the natural feel dimension, these buildings are considered artificial, reflective, smooth, bright, and refined. We will refer to this cluster, which includes 30 buildings, as “expressive design.” Examples include many “blobitectural” buildings, such as the Guggenheim Museum in Bilbao (Spain) designed by Frank Gehry (see Table 2c).

Cluster 3 is differentiated from other clusters by four dimensions—elaborateness, harmony, natural feel, and transparency—with above-average scores. The perception of harmony is particularly high and differs significantly from all other clusters. The design is perceived as harmonious, comforting, and coherent and is clear, elegant, planned, timeless, and protective. The building form is seen as proportional and is also characterized by a natural feel dimension. In contrast to Cluster 2, the buildings in Cluster 3 are perceived as natural in terms of their materials and as rustic and cozy in terms of their secondary design attributes. In terms of the transparency dimension, the buildings are considered open, transparent, weightless, and graceful. We will refer to this cluster, which includes 38 buildings, as “balanced design” because of the well-balanced, harmonious, and overall pleasing design components of the buildings within this cluster. Examples include the Wieshaupt Forum designed by Richard Meier in Schwendi (Germany) (see Table 2c).

Finally, Cluster 4 has below-average elaborateness values. The building design is impersonal, common, unimaginative, banal, ordinary, and monotonous. The building design is considered geometrically bound and technical in its form, with a functional flat façade design; however, the building design is also considered slightly harmonious. The building impressions include a lack of transparency (that is, closed, opaque, weighty, and firm building impressions) and a lack of colorfulness (that is, a bland, discreet, and monochrome color scheme). We will refer to this cluster, which includes 36 buildings, as “solid design.” Contemporary building examples are again shown in Table 2c.

Moreover, while the experts felt that all design types were equally plausible as corporate architecture (M = 5.18–5.53; no significant differences), the experts evaluated balanced and expressive designs (M = 5.28; M = 5.05, respectively) as being significantly more aesthetically appealing than disruptive and solid designs (M = 4.05; M = 4.30, respectively), p < .001. There was no significant difference in the means of disruptive and solid designs or between balanced and expressive designs.

How are the four prototypical architectural designs that emerged related to functionalist and experiential architecture types? It appears that solid and balanced designs are two subtypes of functionalist architecture, whereas expressive and disruptive designs are two subtypes of experiential architecture. This view appears justified based on the visual examination of the buildings and the empirically driven interpretations of the clusters. More importantly, we conducted an additional two-step cluster analysis with a forced two-cluster solution. The first cluster included 93.4 % of the disruptive and expressive architectures; the second cluster included 94.5 % of the balanced and solid architectures.

4.3.4. Relating design types and personality dimensions

We conducted a MANOVA, using the previously identified design clusters as independent variables and the brand personality factor scores as dependent variables. The MANOVA was significant (Wilks’ λ = .314, p < .001). Subsequent ANOVAs (all p values < .05), followed by t-tests comparing each value against the sample mean, indicated that disruptive design was associated with excitement and a lack of competence (see Table 2b). Expressive design was associated with excitement and a lack of competence as well as with stylishness. Balanced design was associated with naturalness. Finally, solid design was associated with lack of excitement, competence, and lack of naturalness. Conversely, post hoc tests conducted row-wise indicated that the brand personality dimensions of stylishness and naturalness were each clearly associated with one architectural design—expressive design and balanced design, respectively. Excitement was strongly associated with expressive design and somewhat with disruptive design (though significantly less), and both expressive and disruptive designs were significantly more “exciting” than solid design, which strongly lacked excitement. Competence was quite strongly associated with solid design (and somewhat with balanced design) and was significantly different from disruptive and expressive design, which lacked competence.

Similar to the experts, consumers also considered all four designs to be equally plausible for company buildings; differences were not significant. Furthermore, similar to the experts, consumers evaluated the balanced and expressive designs as aesthetically more appealing than the disruptive and solid designs (M = 4.81, M = 4.70 versus M = 4.41, M = 4.50). The corresponding contrast was significant at p < .05.

4.4. Discussion of empirical findings

The empirical study identified several relevant design dimensions (elaborateness, harmony, natural feel, and transparency) and design types (solid, balanced, expressive, and disruptive) and showed how they relate to brand personality impressions of excitement, competence, stylishness, and naturalness. The results indicated that there appears to be a trade-off between competence and excitement: if a design was associated with competence, it was not considered exciting, and vice versa. In addition, we presented evidence that expressive and disruptive architectural designs appear to be subtypes of a superordinate experiential architecture, whereas balanced and solid architectural designs are subtypes of a superordinate functionalist architecture. Thus, generally speaking, excitement appears to be associated with experiential architecture, and competence appears to be associated mostly with functionalist architecture.

One might consider functionalist and experiential architectures as a continuum, with solid architecture being the functionalist extreme, followed by balanced design as the moderate version of functionalism, followed by expressive design as the moderate version of experiential architecture, and, finally, followed by disruptive design as the extreme form of experiential architecture. Thus, solid design appears to be starkly functionalist, and disruptive design appears to be excessively experiential. Interestingly, when the design dimensions were used to describe the two extreme designs, the designs were primarily characterized negatively by what they are not: solid design is not elaborate, transparent or colorful and disruptive design is not harmonious. In comparison, the moderate designs were described positively by what they are: balanced design is harmonious, natural, and transparent and expressive design is elaborate.
Both expert architects and non-expert consumers considered the two extreme forms of functionalist and experiential architectures, solid and disruptive designs, to be less aesthetically appealing than the two moderate forms of functionalist and experiential architectures, balanced and expressive designs, resulting in an inverted-U effect of the architecture type on aesthetic appeal. Furthermore, in terms of personality associations, while competence tends to be associated with the two functionalist design subtypes and excitement tend to be associated with the two experiential design subtypes, the two other personality dimensions appear to be associated only with the more appealing design types (naturalness with balanced design and styliness with expressive design). A similar inverted-U effect has been found in research on aesthetics (Berlyne, 1970), where a moderate level of arousal was experienced as being rewarding and pleasant. Furthermore, in prior research on typeface and logo design, moderately compressed typefaces were considered more pleasing and moderately elaborate logos were considered as having a more positive affect than high and low levels of elaboration (Henderson & Cote, 1998; Henderson et al., 2004).

To examine whether the concept of a functionalist-experiential continuum (from solid architecture to balanced design to expressive design and disruptive design) that we proposed here holds true and to further validate the proposed inverted-U effect of architectural design on aesthetic appeal, we conducted a second study with a group of prominent, professional architects.

5. Study 2: qualitative validation of architectural categorizations

5.1. Method

Ten professional architects (nine prominent architects in leadership positions and one city planner) participated in Study 2. Zaltman and Coulter (1995) have shown, across 20 projects, that for qualitative-visual studies, relatively few (average = 5.6) interviewees are needed to generate 90% of the qualitative constructs. We first showed participants groups of five buildings for the four architectural design types identified in Study 1 and asked them to form two superordinate categories. The examples provided to participants have previously been shown in Table 2. Next, we asked participants to articulate the differences between the two superordinate groups that they formed and, subsequently, the differences within each superordinate group particularly focusing on the desired consumer impressions associated with each design group.

5.2. Results and discussion

In creating superordinate categories, all ten architects included expressive and disruptive designs in one group and solid and balanced designs in another group. The descriptions of the superordinate groups mirrored closely the concepts of functionalist architecture (abbreviated as “FU” below) and experiential architecture (abbreviated as “EX” below), as the following representative examples illustrate:

“[EX] buildings are intentional objects, like sculptures, and monuments. [FU] architecture is simply built mass.” (Interview 4)

“[EX] architectures build identity; they have a clear “face”; a very striking layout, which clearly goes beyond function.” (Interview 3)

“[FU] buildings use classic, archetypical forms, such as a wall and a roof. In [EX] architecture, these differences dissolve: the wall is the roof. It’s a hybrid. Moving from [FU] to [EX], the architecture looks more and more like a logo.” (Interview 6)

In addition, respondents’ descriptions of the two groups within each superordinate group confirmed our proposition that there is a continuum, with disruptive and solid designs being polar extremes and expressive and balanced designs as moderate forms. Disruptive design was described as being exaggerated, illogical, disproportional, and arbitrary and therefore was considered to be an extreme form of experiential architecture that is less appealing than expressive architecture. Accordingly, architects expected that expressive design would lead to desirable consumer impressions such as harmony, originality, and dynamics. In contrast, disruptive design was expected to be perceived as cold, egoist, and “harsh.”

“These buildings [disruptive] are arbitrary, design elements are randomly combined. In contrast, these buildings [expressive] are really well designed; there is a strong idea behind them; they are almost monumental and much more harmonious.” (Interview 2)

“[Disruptive] is unique but in a very dictatorial way. It has nothing to do with the product. It’s just [for] show-off. It’s cold and egoistic. [Expressive] is unique and emotional. It’s dynamic. It’s designed for people.” (Interview 10)

Similarly, solid design was described as precise and serious but also as anonymous and reduced. Balanced design, in contrast, was considered to be precise, serious, conservative, and thoughtful but also inviting, open, transparent, and friendly.

“[Solid] is anonymous. (...) It displays solidity but also stagnation. [Balanced] is similar but somehow more humane.” (Interview 3)

“[Solid] is focused, thoughtful, not egocentric. [Balanced] is a little bit more open, more inviting.” (Interview 7)

Moreover, participants drew inferences about the type of organizations and even industries that might use the different designs and their staff, culture, corporate brand image and personality.

“[Solid] seems like architecture for administrative buildings. For those companies that employ cold and unemotional people. It is simply functional. [Balanced] is for pharmaceutical companies. For companies that try to be innovative and to do scientific research. The buildings have a certain originality. Could also be for a publisher or media company.” (Interview 10)

“[Expressive] may be used by a global, innovative player. [Solid] reminds me of companies located in an industrial area.” (Interview 8)

“[Disruptive] seems to be for corporations that want to attract attention to themselves at all cost. This is very showy.” (Interview 7)

6. General discussion

6.1. Toward a consumer-oriented design theory

Examining our research in a broader context, Table 3 presents a consumer-oriented model of architectural design, and Table 4 presents the key similarities of design types across design domains (logos, packaging, typeface, and architecture). The content of the two tables may serve as a basis for developing a consumer-oriented design theory that covers architectural design, as well as other design domains.

As shown in Table 3, starting at the top of the table, out of the individual design attributes of buildings, several design dimensions emerge empirically; these design dimensions differentially characterize architectural types. Solid architectural design appears to be closely associated with the traditional functionalism described in architectural theory, which avoids elaborateness in design, transparency and color but has a fairly harmonic structure. Balanced design appears to result from relaxing some of the original principles of functionalism by focusing on harmony, transparency and a natural feel and even allowing some elaborateness in the design. Expressive design appears to represent the postmodern turn in architecture, focused on elaborateness and multiple allusions. Finally, disruptive design represents contemporary deconstructivist architecture, which is intentionally not at all harmonious. These four types may be viewed
as subtypes of functionalist and experiential architectures. The core brand personality associated with functionalist architecture appears to be competence and, for its later “balanced” variety, naturalness; the core brand personality of experiential architecture is excitement and, for its “expressive” variety, stylishness. Moreover, as mentioned earlier, there appears to be a trade-off between competence and excitement. Although the two experiential design types (expressive and disruptive) were considered exciting, neither was considered to be competent. Similarly, whereas the two functionalist design types (solid and balanced) were considered competent, the “solid” type was considered unexciting and the “balanced” type was considered neutral in terms of excitement. As shown in Table 4, several consumer responses to architectural designs in the present studies are comparable to consumer responses to other visual stimuli found in prior research. In particular, the design dimensions of study 1 share similarities with the design characteristics established in previous research on logo (Henderson & Cote, 1998), typeface (Henderson et al., 2004) and packaging designs (Orth & Malkewitz, 2008). The research within these three design domains established the design dimensions of “elaborateness,” “harmony,” and “natural feel,” as we have done here for architecture. The architectural design factor “transparency” may be considered the equivalent of the design factor “weight” that was identified for typeface and package design. Finally, while “colorfulness” emerged as a separate dimension in the present research, it may be considered as contributing to harmony and naturalness in other visual domains. Alternatively, colorfulness may not be a shared dimension across design dimensions, similar to “flourish” or “compressed,” which emerged only in typeface and packaging design, or to “round,” “proportion” and “repetition,” which emerged only in logo design.

Moreover, as shown in the columns of Table 4, the four architectural design types (disruptive, expressive, balanced, and solid) show considerable overlap with similar types identified in related visual design research. Particularly, based on the visual examples and terminologies shown, in each visual design domain there appear to be functionalist and experiential designs, and even subtypes that show similarities across domains. Finally, the trade-off between competence and excitement found in this research was also observed in some prior design research (Henderson et al., 2004; Orth & Malkewitz, 2008).

Such similarities in visual design perceptions across vastly different domains (from small elements, such as typefaces and logos, to products, packaging and large-scale corporate buildings) raise the question of whether there may be a general design language. This design language may consist of universal design elements related to elements of space, vertical and horizontal lines, and solids and wholes, as discussed in the classic work on visual thinking by the Gestalt psychologist Arneiln (1969). Semiotically, these design elements may form the text and provide the communicative sign language of all visual design. Non-design experts may perceive these elements not individually but collectively along certain common design dimensions and may categorize them into certain design types, of which functionalist and experiential appear to be the most superordinate.

6.2. Future research on consumer behavior and architectural design

The general consumer-oriented design theory that we have described herein needs to be further developed conceptually and empirically tested. First, researchers should explore whether the proposed design dimensions, design types and personality impressions can be cross-validated in design domains that have thus far not been empirically examined, such as in interior design, event design and graphic web design. Second, studies should examine the process through which consumer perceptions of designs occur. For example, do consumer perceptions occur in a sequential fashion (e.g., starting with a rather holistic perception of design elements, which are then perceived along design dimensions, then categorized into design types, and finally matched with personality associations)? Most importantly, research should move beyond the study of individual design domains (typeface, logo, product, packaging, architecture as well as interior spaces, event, and web design) and explore whether consumers perceive visual similarities across domains. Such visual similarities across domains may be a prerequisite for developing a visual style for a brand that facilitates cross-functional integration in corporate communications.

In addition, future research should also move from architecture to interior design. Such an extension would allow the examination of the effect of architecture on employees (Raffelt, Littich, & Meyer, 2011), for example, by examining actual behavioral outcomes of

Table 3
A consumer-oriented architectural design model.

<table>
<thead>
<tr>
<th>Design attributes</th>
<th>Primary (form, façade, material, color) and secondary design attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design dimensions</td>
<td>Harmony</td>
</tr>
<tr>
<td></td>
<td>Elaborateness</td>
</tr>
<tr>
<td>Subtypes of architecture</td>
<td>Solid</td>
</tr>
<tr>
<td>Type of architecture</td>
<td>Functionalist</td>
</tr>
<tr>
<td>Sign language</td>
<td>Denotative</td>
</tr>
<tr>
<td>Brand personality</td>
<td>Competence</td>
</tr>
</tbody>
</table>

Table 4
Design types share similarities in different visual design domains.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Logos</th>
<th>Unproportioned</th>
<th>Elaborate</th>
<th>Harmonious</th>
<th>Not elaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product/Packaging</th>
<th>Contrasting</th>
<th>Delicate</th>
<th>Natural</th>
<th>Massive</th>
<th>Canon</th>
<th>Weighty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unharmonious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elaborate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harmonious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

functionalist and experiential designs in service settings. Functionalist designs, associated with competence, may result in more confident service by employees, and experiential designs, associated with excitement, may lead to a more engaging service.

Several limitations of the present studies should also be addressed in future research. First, the relationship of design and personality dimensions (for example, competence or excitement) may depend on the type of company. A shipping company or bank company might be considered more solid and functional whereas a fashion or entertainment company might generally be more stylish and exciting. Therefore, the effect of product category and industry should be explored in future research. Second, in our studies, the buildings were presented as images and could not be experienced in their full dimensionality and context. Although there is no reason to expect that a three-dimensional experience of the architecture would change the direction of the effects, experiencing the architecture in this manner might affect the strength of the effects because the relevant architectural design elements that evoke associations may be more salient. However, the context of the architecture, including variables such as crowding, proximity to other buildings, and the area in which the building is situated, may change the direction of some effects. For example, functionalist and experiential buildings may be perceived differently in industrial, urban or rural contexts because of the assimilation and contrast effects between the buildings and their immediate surroundings.

Finally, although not explored in prior research (e.g., Aaker, 1997), the personality dimensions of competence and naturalness (or sincerity) associated with functionalist designs appear to be structurally different from the dimensions of excitement and stylishness (or sophistication) that are associated with experiential designs. The former personalities appear to be rather cognitive-analytical or intellectual, whereas the latter personalities appear to be hedonic or sensory-affective in nature (Brakus, Schmitt, & Zaranontello, 2009), which may be an essential, structural difference between these dimensions. Furthermore, another structural difference may be that competence and naturalness is considered as inherently belonging to objects whereas excitement and, to a large degree, styliness appear to refer to an active relationship between the object and the observer (the object provides excitement for the observer; the object provides style and aesthetic appeal for the observer). Thus, future research should examine whether functionalist design may be perceived and evaluated in an analytical, object-centered fashion, whereas experiential design may be perceived in a sensory-affective and relational fashion.

6.3. Managerial implications

The present research provides specific recommendations for management and architects on how to use the architecture of a corporate building to project a specific corporate image and personality. Once management has decided on the desired corporate brand image and personality and is working with an architect on the building project, management should ask architects for functionalist designs if they want the organization to appear as rational and competent and for experiential designs if they want the organization to be considered exciting.

More generally, the present article adds a much-needed multidisciplinary perspective, bringing together the fields of architecture and design and marketing. For a long time, architecture and marketing have developed as separate disciplines, and there has been little communication between them (Ostrander, 1974). The current research bridges these two disciplines by presenting consumer-centric and marketing-based views of architectural designs. Our research may thus be viewed not only as a conceptual and empirical model but also as an inventory that architects may use to better communicate with managers about design. Conversely, managers can use the inventory to better communicate the intended personality of their corporations to architects. In addition, our research may be useful for city planners who use architecture to position entire cities or city districts, as many Asian and Gulf State cities as well as European (e.g., Bilbao) and U.S. cities (e.g., Baltimore) have done recently. To be sure, image or personality is only one criterion for planning a project for a city; other key considerations in such ventures may include financial and opportunity costs, as well as the sustainability of the project. We hope that the future corporate or public buildings, which will result from collaborative processes between architects, designers and marketer, will use form and function more effectively to create the desired corporate and civic personalities for consumers and citizens.

References


Photo credits

BMW Tower, Munich, Germany: courtesy of BMW Group.

BMW World, Munich, Germany: courtesy of BMW Group.

Bugatti Manufactory, Molsheim, France: courtesy of Henn Architekten.

Businessparc, Cologne, Germany: courtesy of Kruse GmbH, Germany.


China Television (2008), Peking, China: © Ou Bal.


Columbia Business School (1961/1984), New York City, USA: own picture.

Columbia Business School (planned), New York City, USA: © UNSStudio.

Congresso Nacional, Brasilia: © Christof Mauersberger.

Freitag Flagshipstore, Zurich, Switzerland: courtesy of Freitag GmbH, Zurich, Switzerland.


High Court of Justice and Supreme Court, Madrid, Spain: courtesy of Foster + Partner, London, United Kingdom.

Mercedes-Benz Museum, Stuttgart, Germany: courtesy of Daimler AG.

Peak Tower, Hong Kong, China: Wing, “The Peak Tower,” August 14th, 2011 via Wikicommons, Creative Commons Attribution.

Prisma Research Building, Lustenau, Austria: courtesy of Prisma GmbH, Austria.


Sampselion, Hellerup, Denmark: © 3XN / Adam Moerk.


Strata Tower, Abu Dhabi, United Arab Emirates: courtesy of Hani Rashid + Lise Anne Couture.

Telus Pavillon, Montréal, Canada: © Paul Lowry, “TELUS Pavilion on the UQAM Campus,” December 16th, 2008 via Flickr, Creative Commons Attribution.


The Wave, Vejle, Denmark: Photo by Thomas Moelvig.

Weishaupt Forum, Schwendi, Germany: courtesy of Max Weishaupt GmbH.

Zenith Music Venue, Saint-Etienne, France: courtesy of Nigel Young/Foster + Partners.
When white space is more than “burning money”: Economic signaling meets visual commercial rhetoric

John W. Pracejus a,*, Thomas C. O’Guinn b,1, G. Douglas Olsen c,2

a 3-40J Business Building, University of Alberta, Edmonton, Alberta T6G2R5, Canada
b University of Wisconsin, Madison, 975 University Avenue, WI 53705, USA
c Arizona State University, W. P. Carey School of Business, P.O. Box 874106, Tempe, AZ 85287-4106, USA

A R T I C L E   I N F O

Article history:
First received in 5, January 2010 and was under review for 4 months
Available online 6 April 2013
Area Editor: Amitava Chattopadhyay

Keywords:
Burning money, Design, Visual persuasion, Branding, Persuasion knowledge, Social history

A B S T R A C T

Previous work has demonstrated that the use of white space in advertising communicates specific meanings to consumers and that this meaning derives from particular historical moments in the art and visual rhetoric of 20th-century North America. The use of “empty space” in ads, however, can also be conceptualized as a signal of burning money, which could influence consumer perceptions about the size and power of a company through completely different mechanisms. Somewhat surprisingly, nearly all empirical demonstrations of burning money in a consumer advertising context also manipulated white space, leaving the mechanism of action unclear. The results of the three studies discussed here indicate that white space is different from other ways of burning money, and its meanings are sufficiently different across cultures, thus providing stronger support for the rhetorical explanation than the economic signaling one. The findings are discussed in terms of their implications for previous research that found that consumers infer quality from the economic signals of burning money.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The pioneering work of Nelson (1970, 1974, 1978) introduced the idea that, under certain conditions, high advertising expenditures can inform consumers about particularly high product quality, even if consumers never actually see any of the advertising. That is, the mere existence and expense of the ads, not the information in the ads, will produce a positive aggregate consumer response. Presumably, this effect occurs at the individual level due to knowledge, however vague and sparse, of the ad’s existence and prevalence. Large advertisers, i.e., those with “money to burn,” spend generously. Consumers casually notice the rapid combustion of these big spenders’ ad dollars. This presumed, vague awareness, independent of the actual persuasive information (e.g., copy points, claims, images) within the ads themselves, leads to consumer perceptions of higher product quality. This intriguing assertion, which is often referred to as “burning money,” has subsequently fuelled economic theory and empirical investigation into when and how this and similar “signals” affect consumer perceptions of product quality (Kihlstrom & Riordan, 1984; Schmalensee, 1978; Zhao, 2000). In 1986, Milgrom and Roberts (Milgrom and Roberts, 1986) restated the “burning money” idea:

It is clear that if high quality brands advertise more and if advertising expenditures are observable (even if not perfectly so), then rational, informed consumers will respond positively to advertising, even if the ads cannot and do not have much information content. (p. 797)

This begs several questions, including the following: Does the “burning money” effect actually exist? What behavioral evidence is there for this effect? What individual level theory could best explain it? What is “information content”? This paper offers a closer look at an effect that is sometimes attributed to “burning money,” proffers a different theoretical explanation, and draws attention to the problematic ways in which different traditions materially investigate “information.”

2. Theory

2.1. “Burnning money”

From 1974 until 1989, all papers addressing the topic of “burning money” were either conceptual/theoretical treatments or analytical modeling papers, or both. The first attempt to empirically demonstrate that consumers actually do what the economic models predict was by Kirmani and Wright (1989), who demonstrated that...
providing (or cuing) actual advertising expenditures can influence consumer brand quality perceptions. Of course, no consumers, other than marketing and advertising professionals, are ever presented with this type of tabulated expenditure data. Speaking to this issue, and the questionable real-world evidence for quality perceptions through “burning money,” Becker and Murphy (1993) noted that if this type of signaling really worked, “companies should advertise how much they spend on advertising, yet almost no companies do that” (p. 994). More recent models actually predict that modest rather than large advertising expenditures may signal high quality (Orazch, Overgaard, & Tauman, 2002). Zhao (2000) argues that when advertisers move from pure signaling to more “informational” (p. 390) goals such as awareness (p. 395), “burning money” will actually be detrimental to the higher quality firm due to increased mimicry by lower quality brands. It should be noted how “signaling” is distinguished from “information” within this literature.

Very few studies have produced evidence that advertising stimuli, as opposed to telling consumers how much was spent on advertising, can actually impact quality inferences in consumers. Two studies (Kirmani, 1997; Moorthy & Hawkins, 2005) manipulated advertising frequency. As advertising frequency is almost always highly correlated with the total advertising expenditure, these studies offer some support for the “burning money” proposition. However, as Moorthy and Hawkins (2005) noted, ad frequency effects are thoroughly confounded with mere exposure effects, leaving the role of expenditure signaling in doubt. Two other studies (Homer, 1995; Kirmani, 1990), both by consumer psychologists, manipulated advertising size. In these studies, the relevant “burning money” idea is that ad size is significantly related to the size of the ad spend and company size. Large ads are more often placed by large and generally more well-known (trusted) advertisers, whereas small ads usually come from smaller, and less well-known (untrusted) advertisers. Kirmani (1990) found that ad size affected perceived costs, which in turn affected brand perceptions. Ad size, however, did not affect brand perceptions directly. This was most likely due to noise in the model (p. 170). In this mediated effect, medium-sized ads provided the most positive brand quality perceptions, with the small (one quarter page) and the very large (eight pages) ads having lower quality perceptions. Homer (1995) reports a main effect of ad size on quality perceptions, but this effect was mediated by a perception of the larger ad being better designed.

Like the economic work, these two papers give limited support to the “burning money” idea. They are similar to economic work in that perceptions of higher product quality are produced by either a surrogate measure of advertiser size, the size of the ads themselves, or, in the case of Kirmani (1990), the effect of the perceived cost of the ads. The effect was never said to be due to what would typically be called “information” residing within the ads, i.e., body copy, headlines, verbal arguments or even pictorial representations or demonstrations. This effect, however, is not observed in the case of very large ads. In Kirmani (1990), eight-page print ads lead to lower brand quality perceptions. This inverted U-shaped relationship is also observed in Homer (1995) but only through the use of directionality tests of regression coefficients. Homer’s (1995) study had only two conditions of ad size: large and small.

Unfortunately, both Kirmani (1990) and Homer (1995) confounded white space with advertising size. Kirmani (1990) stated:

> Because the text and picture were kept constant across ads, the longer ads had more white space than the smaller ads. Keeping the elements and copy constant insured that there were no informational differences that might account for effects on brand perceptions (p. 65).

Homer (1995) also manipulated white space, but not in a systematic manner. She states, All information and ad execution elements (layout, typeface, etc.) remained constant across ad size treatments. Typographical information was reduced or enlarged (depending on ad placement) to project a professional appearance consistent with the original ads (p.4).

It is therefore impossible to know what the effect of white space was on brand quality perceptions. In the manipulation checks, Homer (1995) reports,

> Respondents did rate the larger ad (mean = 4.93) higher in terms of the ‘design’ dimension of ad cognitions (F(1,240) = 8.81, p < 0.05) compared to the smaller ad (mean = 4.38), which is understandable due to the less cluttered appearance resulting from the additional white space.... (p. 6).

Again, ad size and white space were confounded. Unaware of the rhetorical theory of white space that would later be developed by Pracejus, Olsen, and O’Guinn (2006), both Kirmani (1990) and Homer (1995) mentioned the white space issue, but did not consider the possibility of its central role in a completely different explanation for their findings. This omission is particularly critical given the above mentioned confound between mere exposure and signaling in frequency studies.

White space is a well-known design element used in advertising. Since its first use in the 1920s, advertisers have often fought their advertising agencies over its application, claiming that it was wasteful, akin to burning money. During the very frugal 1930s, white space almost disappeared because advertisers demanded every paying inch of an ad (Fox, 1985; Marchand, 1985) be filled with words and multiple pictures to justify their media costs. During the Great Depression the idea of paying for “empty space,” (i.e., white space) was occasionally said to be like “burning money.” This conflation makes it easy to see how white space is connected to “burning money.”

### 2.2. White space as trope

To this point, the “burning money” proposition had been theorized by economists and empirically examined by either economists or consumer psychologists. While silent on the “burning money” question itself, Pracejus et al. (2006) offer an entirely different theoretical explanation of the effect of white space on consumers. They found that white space was related to significantly higher levels of brand quality, prestige, leadership, trust, as well as lower levels of risk. Thus, this study that more deliberately isolated white space found a series of related positive effects of this “burning money” variable. Most significantly, Pracejus et al. (2006) theorize white space as neither a narrow economic signal nor a peripheral cue (Homer, 1995). Instead, they theorize white space as a part of commercial speech, a trope of visual rhetoric. Pracejus et al. (2006) argue that white space is a well-known and widely understood visual trope in the North American advertising lexicon, and one that produces the same “burning money” effect of higher product quality but also a constellation of other related meanings. They argue that members of a given consumer society (in this case, North America) have learned to “speak” the advertising language and that white space is part of that visual lexicon.

This is significantly different from the elaboration-reliant consumer-information-processing work of both Kirmani (1990) and Homer (1995). Members socialized in a consumer society in which white space has certain meanings, know what an ad with a lot of white space means. Unlike the other behavioral works, the recognition and understanding of this visual language is not accompanied by explicit assumptions and attributions about the advertiser, nor is white space relegated to the affective world of “peripheral cues.” The Pracejus et al. (2006) explanation is linguistic, rhetorical and socio-historical. They posit these responses as requiring only the near-automatic recognition and
subsequent understanding of a well-known visual trope, not unlike common parts of speech are understood.

Pracejus et al. (2006) traced the emergence of white space in North America to mid-20th century minimalist art and to its appropriation by early designers of corporate images and advertising professionals seeking to convey trustworthiness and high quality. White space owes a significant amount of its present meaning to its connection to this minimalist movement in art and architecture, corporate design, and then advertising. Many of its artists, including Robert Rauschenberg, were very interested in materiality, including mass materiality. The movement’s architects, such as Mies van der Rohe, argued that less is more and tied “empty” spaces to aspirations of upward mobility (Sennett, 1990). They link white space to visual representations of the “clean” spaces of upscale consumers in contrast to the dense and cluttered spaces of the urban poor. Corporate designers, such as Paul Rand, made minimalist corporate images for leading companies such as IBM. The enormously important creative revolution of the 1960s further institutionalized the meaning of white space in advertising. The fact that all of these streams of cultural production came together at the same time and place to give white space a particular meaning is entirely relevant here. It is history that gave white space its meaning, not mere affect, or a narrow economic signal. White space is not just “burning money,” at least not in North America.

Even though white space has long been connected to “burning money,” Pracejus et al. (2006) make the argument that white space is much more than what the narrow term “signal” implies; it is also a lot more than a “peripheral cue” that works through affect. In fact, there is nothing peripheral about it. Instead, white space is an understood visual trope with meaning much richer than “burning money” or high product quality. It is a linguistically stable part of contemporary American commercial speech with a decades-long social history. These authors argued that white space means what it does because of this particular history. This meaning does not require a lay understanding of advertising expenditures, nor elaborations and attributions about the nature of the company via a “schemer schema” (Wright, 1986). The meaning of white space derives from a member of a particular consumer society simply being literate in the visual language of advertising.

Although Pracejus et al. (2006) supported their account through two studies, neither of which directly addressed “burning money” as an alternative explanation. This was unfortunate because “burning money” is the leading alternative explanation for their findings. Further, given that white space was fully confounded with ad size in Kirmani’s (1990) study and in an unknowable way in Homer’s (1995) study, it is both necessary and important to more directly contrast the competing economic and socio-historical accounts. Do history, culture and language add to our understanding of the previously reported effects of ad size on brand quality inferences compared to economic signaling alone? Does ad size net of white space impact quality inferences? Does white space have certain meanings, compared white space against the only candidate for an advertising element that might signal product quality: large ad size. Given the previous confounding of these two constructs, we used a unique experimental design that allowed us to parse out the two mechanisms.

Investigating the range of meanings evoked by white space compared to ad size should shed light on how white space communicates. Is it just like ad size: a narrow form of communication, or signal, through which the consumer is thought to make the inference: “this company has so much money they can just ‘burn it’ with big ads and empty (or white) space.” Conversely, does white space evoke a richer and more nuanced meaning? Stated differently, is it information of a linguistic sort? Does white space have certain meanings, and if so, how rich are those meanings?

The second issue, i.e., cross-cultural embeddedness vs. consistency, is explored in Studies 2 and 3. If economic signaling fully explains consumer quality inferences, there should be little, if any, variance in meaning across cultures. If the rhetorical and linguistic account is valid, then there should be a different array of evoked meanings in North America than in cultures with a different history.

3. Study 1

To more fully contrast the rhetorical account of white space with the economic “burning money” account, we first compared white space against ad size, which is something Pracejus et al. (2006) did not report. These authors also argued that rather than a simple narrow signal, the specific history of the white space trope gives it a much fuller meaning, independent of ad size.

To test this, we started with a small (1/5th page) ad that does not use white space as a design element, and compared it with a large ad (full page), also without white space, and a large ad (full page) with substantial white space. Note that the two full-page ads would cost the same, and would serve as an indicator of equivalent cost to consumers. Investigating the role of white space independent of ad size presented some uniquely challenging experimental design problems. It is not possible to manipulate white space while holding both the physical size of the image and ad size constant (i.e., it is not possible to hold image size constant and manipulate ad size without also manipulating white space). Because of this, a standard full factorial design (i.e., ad size by white space by image size) is not possible. This apparent conundrum was solved by employing a series of planned contrasts between the small ad and each of the two full-page ads. This allowed the impact of ad size, alone, to be compared with the additional effect of white space. Although this was an extremely conservative test,3 it was the only one available given the abovementioned constraints. Two different product replicates, mutual funds and furniture, served to demonstrate robustness across categories.

3.1. Participants

One hundred and eighty-four undergraduate students at a North American university participated in this study for research credit.

3.2. Design and stimuli

Three different ad conditions (full-page, low white space; full-page, high white space; and 1/5th page, low white space) were used, with 2 product category replicates for each ad (furniture and mutual funds).

The advertisements were printed on 11-inch by 14.5-inch paper (the approximate size of a tabloid newspaper). The cover story for the study was that we were interested in international advertising. Participants

3 The test is conservative in that it requires planned contrasts between 1/5th page and 1 page, high white space for variables that do not show significant contrasts between the 1/5th page ad and the full page, low white space ad (i.e. a very specific pattern of results).
were told they would view an actual ad for Hastings (furniture/clothing) that was about to be run in The Guardian, a British newspaper. In the 1/5th page ad condition, the ad was placed in the lower right hand corner of the page, surrounded by 4 columns of nonsense text, to provide the same level of contrast that might be expected in a newspaper reading environment (see Fig. 1). All advertisements were viewed in a lab where individuals were unable to see what other participants were reading.

The full-page ads were 10 by 12.75 in. The image in the full-page, low white space ad was 9.75 by 9.75 in., whereas the high white space version had a 3.5 by 3.5-inch image. The fifth-page advertisement was 5 by 6.5 in. in dimension and employed an image that was 3.5 by 3.5 in. A fictitious brand (Hastings) was used. The copy was designed to sound realistic:

Sit for a while. No matter what day of the week or week of the year, there are many things that we have to deal with. From what to wear...to where to invest [buy furniture]. We don't pretend to have the answers to all of life's demands, but we can help when it comes to mutual funds [furniture]. In a world where every day is different you can depend on us to give you consistency and comfort. Visit us on the web at www.hastings.com or call 1-800-686-3321 today to find out where our mutual funds are [furniture] is available.

3.3. Dependent variables

After viewing the ad for 30 s, participants were asked to rate their level of agreement with 10 statements about Hastings (see Table 1) using a seven-point scale (1 = “Strongly Disagree” and 7 = “Strongly Agree”). Eight of these statements were from Pracejus et al. (2006).

We were also interested in testing the possibility that the findings of that paper were based on “burning money,” leading to market power perceptions followed by a halo effect, and causing spill-over to other measures such as trust and expensiveness. To test for this possibility we included two yeah-saying foils that the social history would not predict to be affected by white space but that might be susceptible to halo effects spilling over from market power. These were: Hastings promotes equality in the workplace and Hastings donates considerable money to the community. Attitudes towards Hastings and purchase intentions, if Hastings becomes available, were also taken.

3.4. Results

An initial MANOVA was used to assess the dependent variables, using the product category and nature of ad as independent variables (see Table 1). The product category demonstrated a number of main effects, with the furniture product category demonstrating higher values on the quality, not risky, expensiveness, general attitude toward the brand and purchase intentions. Product category, however, did not interact with the type of the advertisement.

For cases where the type of advertisement differed, contrasts were conducted to examine the source of the difference. The impact of ad size alone can be assessed through contrasting the 1/5th page ad with the full page, low white space ad. Here, the only significant impact of ad size is on the item Hastings has considerable market share (F(1,119) = 14.71, p < 0.001). This is in keeping with the expectations of the “burning money” signaling theory.

When the 1/5th page ad was compared with the full page, high white space ad, however, we observed significant increases in beliefs about Hastings in 6 out of the 8 specific meanings expected to be conveyed by white space, as well as significant increases in attitudes and purchase intentions (see Table 2). Finally, it should be noted that neither ad size nor white space had any impact on the two foil items, making a halo effect an unlikely explanation for the impact of white space on perceptions.

3.5. Discussion

It appears that the use of white space in ads had a very different impact on consumers than merely employing equally expensive ads of the same large size. Employing ads with white space increased perceptions about quality, leadership, trust and prestige. Employing equivalently large ads without this design element did none of these things. It appears, therefore, that the impact of ad size net of white space is only on perceptions of large market share, not on quality or the other variables investigated. This provides evidence that white space is more than “burning money” and that ad size, as a candidate for signaling quality, is questionable. Whereas large ads alone boosted beliefs about market share, only the white space ads boosted beliefs about other specific meanings, including quality. Further, these meanings are completely consistent with the trope’s social history. While ad size signaled market share size, ad size without white space did not evoke quality.

4. Study 2

The second way the rhetorical account can be distinguished from the “burning money” account is cross culturally. The rhetorical explanation is a linguistic one; the meanings conveyed through white space are derived from a particular North American social history. Because vocabularies, including visual ones, are significantly a product of specific histories, we would not expect the same meaning to be conveyed by the same trope without a shared history. Pracejus et al. (2006) claimed...
that the white space trope gained its meaning from its historical origins in the mid-20th century minimalist art movement and its migration into advertising and other commercial rhetoric in the decades that followed. Their account claimed that white space means what it does because of its very specific and culturally bound North American social history. Given this, if the socio-historical account is accurate, we should expect a different—weaker, narrower—impact of white space on consumers outside of the North American context.

To explore the role of culture and social history on the meanings conveyed to consumers, we conducted a study involving a white space ad to indicate quality, prestige, trust, and overall brand attitude. The interaction was also marginally significant on expensiveness ($F(1,256) = 3.4, p > 0.06$). No interaction effect was observed on either of the foil variables, which were included to test for yaah-saying (Donates to community, Equal opportunity employer). Therefore, for nearly every meaning associated with the social history of white space, its impact was primarily dependent on culture.

### 4.3 Results

Table 3 reports mean values by condition. It appears that the impact of white space on brand perceptions varied significantly by country. After mean centering by country and product class, an ANOVA revealed a significant interaction of white space and country of residence on quality, risk, prestige, trust and overall brand attitude. The interaction was also marginally significant on expensiveness ($F(1,256) = 3.7, p > 0.055$) and leadership ($F(1,256) = 3.4, p > 0.06$). No interaction effect was observed on either of the foil variables, which were included to test for yaah-saying (Donates to community, Equal opportunity employer). Therefore, for nearly every meaning associated with the social history of white space, its impact was principally dependent on culture.

Table 3 reports mean values by condition. It appears that the impact of white space on brand perceptions varied significantly by country. After mean centering by country and product class, an ANOVA revealed a significant interaction of white space and country of residence on quality, risk, prestige, trust and overall brand attitude. The interaction was also marginally significant on expensiveness ($F(1,256) = 3.7, p > 0.055$) and leadership ($F(1,256) = 3.4, p > 0.06$). No interaction effect was observed on either of the foil variables, which were included to test for yaah-saying (Donates to community, Equal opportunity employer). Therefore, for nearly every meaning associated with the social history of white space, its impact was principally dependent on culture.

Follow-up analysis revealed that, for the North American participants, white space significantly increased perceptions of quality, low risk, prestige, trust and leadership. It also significantly increased overall brand attitude and purchase intention. For the Hong Kong participants, however, white space only significantly influenced perceived market share ($F(1,128) = 5.29, p < 0.023$). It also marginally impacted leadership ($F(1,128) = 3.57, p < 0.061$) and trust ($F(1,128) = 3.25, p > 0.074$). However, for both of these measures, the impact of white space was the opposite from the North American participants. That is, low white space was marginally predictive of the brand being expensive and trustworthy.

### 4.4 Discussion

It is clear that the two groups of participants responded very differently to white space. Whereas North American participants found the impact of white space in an ad to indicate quality, low risk, prestige, trust and leadership, in Hong Kong it simply meant large market share. Large share is, of course, exactly what we found ad size to indicate in the North American context.

### 4.1 Participants

Two hundred and sixty undergraduate students participated for research credit: 130 at a university in North America, and 130 at a university in Hong Kong.

### 4.2 Design and stimuli

A $2 (product category: mutual funds/clothing) \times 2$ (high/low white space) \times 2 (country of residence) completely between-subjects design was employed. All the advertisements contained a picture of a clock, accompanied by a logo for a fictitious brand (Hastings), along with ambiguous copy like that used for Study 1.

All advertisements were presented on an 8.5-inch by 11-inch sheet of white paper. In the low white space ad, a picture of a square clock measuring 8 in. by 8 in. was present. In the high white space condition the clock was 2 in. by 2 in. The font size of the text information remained constant across conditions. A fictitious logo for Hastings was presented in the lower right hand corner. As in Study 1, individuals were given thirty seconds to examine the advertisement. The participants then responded to the same 12 items using the same scale in Study 1.

### Table 1

Effect of ad size and white space on brand perceptions (Study 1).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mean evaluation by condition</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Furniture (n = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ad (A)</td>
</tr>
<tr>
<td></td>
<td>Low WS (n = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Product (P)</td>
</tr>
<tr>
<td></td>
<td>High WS (n = 32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A × P interaction</td>
</tr>
<tr>
<td>Large company</td>
<td>4.53 (1.25)</td>
<td>4.45 (1.24)</td>
<td>4.53 (1.31)</td>
<td>4.58 (1.25)</td>
<td>4.80 (1.45)</td>
<td>4.03 (1.43)</td>
</tr>
<tr>
<td>Promotes equality</td>
<td>4.03 (0.50)</td>
<td>4.00 (0.58)</td>
<td>4.17 (0.592)</td>
<td>4.03 (0.315)</td>
<td>4.17 (0.950)</td>
<td>4.00 (1.05)</td>
</tr>
<tr>
<td>High quality</td>
<td>4.57 (1.45)</td>
<td>5.38 (1.01)</td>
<td>4.93 (1.04)</td>
<td>4.32 (0.871)</td>
<td>5.27 (1.05)</td>
<td>4.23 (1.30)</td>
</tr>
<tr>
<td>Not risky to purchase</td>
<td>4.83 (0.986)</td>
<td>5.53 (0.950)</td>
<td>4.73 (1.17)</td>
<td>4.23 (1.38)</td>
<td>4.90 (1.30)</td>
<td>4.07 (1.28)</td>
</tr>
<tr>
<td>Prestigious</td>
<td>4.47 (1.14)</td>
<td>5.03 (1.33)</td>
<td>4.13 (1.25)</td>
<td>4.26 (0.893)</td>
<td>4.63 (1.45)</td>
<td>3.83 (1.39)</td>
</tr>
<tr>
<td>Market share</td>
<td>4.43 (1.04)</td>
<td>4.28 (1.11)</td>
<td>3.57 (1.07)</td>
<td>4.13 (0.957)</td>
<td>4.10 (1.27)</td>
<td>3.63 (0.809)</td>
</tr>
<tr>
<td>Donates money</td>
<td>3.50 (1.14)</td>
<td>3.78 (0.792)</td>
<td>3.70 (0.877)</td>
<td>3.90 (0.700)</td>
<td>3.90 (1.35)</td>
<td>3.77 (0.935)</td>
</tr>
<tr>
<td>Expensive</td>
<td>4.43 (1.27)</td>
<td>4.59 (1.32)</td>
<td>4.17 (1.23)</td>
<td>3.61 (0.919)</td>
<td>4.00 (1.26)</td>
<td>3.93 (0.640)</td>
</tr>
<tr>
<td>Can be trusted</td>
<td>4.60 (1.07)</td>
<td>4.94 (0.878)</td>
<td>4.20 (1.27)</td>
<td>4.35 (1.02)</td>
<td>5.37 (1.10)</td>
<td>4.33 (1.06)</td>
</tr>
<tr>
<td>Is a leader rather</td>
<td>4.40 (1.61)</td>
<td>5.22 (1.18)</td>
<td>4.03 (1.47)</td>
<td>4.06 (1.31)</td>
<td>4.77 (1.50)</td>
<td>3.83 (1.34)</td>
</tr>
<tr>
<td>Positive attitude</td>
<td>4.50 (1.30)</td>
<td>5.31 (1.3)</td>
<td>4.43 (1.25)</td>
<td>4.42 (1.03)</td>
<td>4.93 (1.43)</td>
<td>4.07 (1.41)</td>
</tr>
<tr>
<td>Purchase intention</td>
<td>4.03 (0.964)</td>
<td>4.03 (1.09)</td>
<td>3.57 (1.04)</td>
<td>3.29 (1.13)</td>
<td>3.77 (1.04)</td>
<td>3.37 (1.09)</td>
</tr>
</tbody>
</table>

- Mean of seven point scale anchored by 1 = "Strongly Disagree" and 7 = "Strongly Agree".
- * p < 0.05.
- ** p < 0.01.
- *** p < 0.001.

* Mean centering is often used in cross-cultural research to reduce the effect of scale differences (see Van de Vijver & Leung, 1997 for a discussion).
imply in Study 1. So, when we leave the North American context, away from its commercial and historical circumstance that gave white space its social meaning, its impact on brand meaning (large share) can be explained by “burning money.” In the absence of the social history that gave the trope its particular meanings, the rational assumption of a large ad, meaning that a company has money to burn, appears operative. As in Study 1, the impact of “burning money” was only on perceptions of share size and not on perceptions of quality.

While these differences between cultures certainly argue for the social history explanation and against the “burning money” one, a skeptical reader might assert that because this was a quasi-experiment (people were not assigned to country) that a reduced number of dependent variables impacted by white space in Hong Kong represent a null effect, which could be due to any number of factors. While we see directional evidence that white space actually leads to perceptions of lower expense and lower trust, these effects did not reach traditionally accepted levels of statistical significance. To address these concerns, we sought to replicate the observed negative effect of white space with a sample size powerful enough to unquestionably detect the effect. We also sought a country with a different visual aesthetic history from both North America and Hong Kong.

5.8. Study 3

To demonstrate that whether white space is a positive or a negative depends upon culture, we sought a country with an aesthetic social history where “more” seems to be more. India was thought to be a likely candidate, as an intricate and ornate visual aesthetic is strongly represented in many areas of visual representation there (Blacker, 1922). We hoped that with a sufficiently powerful sample, we might detect negative effects of white space on brand perceptions that would be wholly incompatible with a pure “burning money” account.

5.1. Participants

One hundred sixty-nine participants who listed India as their country of residence on Amazon.com’s m-turk service were recruited to participate. Twenty-seven (16%) participants were removed for providing multiple responses from the same IP address and 10 respondents (5.3%) were removed for reporting that they were currently living outside of India, and three participants did both, leaving a usable sample of 135 participants.

5.2. Design and stimuli

A 2 level (high/low white space) completely between-subjects design was employed. The advertisements presented were identical to those used in Study 2, except that they were presented on a computer screen, rather than on paper, and only the mutual fund product class was used. The advertisement was shown on the screen for 30 s after which the participants responded to the white space-related brand perception measures employed in Study 2 (the yeah-saying foils were omitted).

5.3. Results

In our Indian sample, white space had a very different impact on brand perceptions than in North America. On most of the items, there was no significant difference between the high and low white space groups. There was a significant impact of white space on both Company is a leader rather than a follower ($F\left(1,133\right) = 14.71, p < 0.001$) and purchase intention ($F\left(1,133\right) = 4.62, p < 0.033$). For both of these variables, however, white space had the opposite effect compared to what was observed in North America. White space led to significantly lower perceptions of company leadership (4.72/7 vs. 5.17/7) and significantly lower intention to purchase the brand (4.89/7 vs. 5.38/7). For our Indian sample, therefore, it appears that “more is more.”

6. General discussion

The “burning money” idea remains provocative and intriguing. The idea that only a vague sense of the existence and prevalence of a company’s ads can produce the belief that the advertised brand is of high quality seems at odds with consumer behavior research and its decades-long focus on information within the ads. In this instance, the contrast between economic and consumer psychological thought is sharp. Yet, a handful of behavioral studies appear to support a “burning money” effect. While the studies that rely on frequency could be due to mere exposure effects, at least two others (Homer, 1995; Kirmani, 1990) find either main effects for ad size (Homer, 1995) or a mediated effect involving perceived spending levels on brand trust. A third study (Pracejus et al., 2006) deals directly with two “burning money”
variables: ad size and white space, the latter of which has been cast as potentially wasteful and evidence of “burning money” since the 1920s (Marchand, 1985; Fox, 1985). However, limitations in the Pracejus et al. (2006) paper made it an incomplete test of the “burning money” principle.

The studies reported here, such as the Pracejus et al. (2006) study, rely on a fundamentally different theory that is neither economic nor reliant on the idea that both pictures, and in this case, the absence of pictures, were “peripheral” cues. In fact, the linguistic and socio-historical theory presented here explicitly rejects relegating visual tropes to the periphery but instead argues they are, in fact, well-understood parts of a commercial language.

Our data show that white space in an ad can imply that the advertiser is a large company much like a large ad does. In North America, white space in ads means much more: high quality, prestige, trust, etc. Yet in Hong Kong, separated from its social history in North American consumer culture, white space only implied a large market share. In India, white space had no positive impact on brand perceptions and actually had a negative impact on purchase intention. This is convincing evidence that the linguistic account adds substantially to our understanding compared to economic signaling alone. Our results demonstrate of this were critical to Nelson’s (1970, 1974, 1978) model being adopted as prescriptive reality for marketers, we believe our findings on this point alone make a reasonable contribution to the field of marketing. Furthermore, our results challenge some widely held and published views of what information is and is not, particularly where visuals are concerned. Across three traditions—economic, information processing, and the linguistic and socio-historical (Pracejus et al., 2006)—we see very different conceptions of the nature of information, from signal to peripheral cue to language. These are not, however, three names for the same thing, as underlying them are very different theories. The results also underscore the need for more research on the visual aspects of market communication. Words have been privileged in consumer research relative to the declining real world prevalence in advertising. This chasm between practice and academic research existed largely because researchers believed they knew more about how to study words (for a refreshing exception, see Schroeder, 2002). Visuals have been stubbornly problematic for many fields. Finally, the socio-historical theory of visual rhetoric (Pracejus et al., 2006) on which our work is premised is offered as a legitimate alternative to both economic signaling and consumer information processing theories of advertising and its effects.

As for the ads-as-information tradition, we hold that ads contain information and that information is sometimes transmitted, received, and interpreted in a manner consistent with the communicative intent of the advertiser. This even occurs in the case of some “burning money” techniques. Large ads may mean bigger companies (Pracejus et al., 2006). Heavy spending on advertising may even influence stock market valuations through the “actual exposure” (Joshi & Hansson, 2010 p. 27) of investors to advertising. Large ads and high expenditure may be closer, due to a narrow range of meanings, to something an economist would term a “signal.”

The white space trope, however, produced a constellation of theoretically supported meanings indicative of a language part, not a simple signal. From our perspective, advertising is fundamentally rhetorical and linguistic. Ads try to persuade, but not necessarily in some highly involved and elaborated manner. It is quite clear that the meaning of white space cannot be explained by signaling alone when this linguistic device is grounded in a meaningful social history. Furthermore, everyday language acts do not require any type of expert knowledge at all. One does not need to know a thing about the size of relative ad spends or know how white space took on its meaning. Many, if not most, tropes do not require the speaker or receiver to know their linguistic origins, just how to use them appropriately. For scholars, however, knowing how blocks of “information-less nothing” (white space) came to have a rich set of related meanings to consumers gives a credible alternative explanation to the economic notion of “burning money.” An example of the historical development of a word, the simplest verbal element, may be illustrative of this point.

Many may be familiar with the term “hat trick,” which today can refer to any three successes, often in a sports context. Few, however, know the origin of the term. According to the Oxford English Dictionary, hat trick originated in the 19th century, referring to a cricketer who

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mean evaluation by condition</th>
<th>WS × country F(1,256)</th>
<th>p-*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hastings is a large company.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low WS 4.50 (0.965) 4.42 (1.00)</td>
<td>4.41 (1.41) 4.44 (1.44)</td>
<td>1.097 –</td>
</tr>
<tr>
<td></td>
<td>High WS 3.48 (1.23) 3.69 (1.13)</td>
<td>3.94 (0.559) 3.91 (0.854)</td>
<td>0.871 –</td>
</tr>
<tr>
<td>Hastings promotes equality in the work-place.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low WS 4.33 (1.01) 4.22 (1.03)</td>
<td>4.47 (1.18) 5.06 (1.11)</td>
<td>6.86 .005</td>
</tr>
<tr>
<td></td>
<td>High WS 4.03 (1.21) 3.91 (1.00)</td>
<td>4.31 (1.19) 4.89 (1.24)</td>
<td>6.26 .013</td>
</tr>
<tr>
<td>Hastings furniture is/mutual funds are of high quality.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low WS 4.09 (1.06) 4.14 (1.04)</td>
<td>4.06 (1.47) 4.79 (1.25)</td>
<td>4.98 .026</td>
</tr>
<tr>
<td>Hastings furniture has considerable market share.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low WS 3.76 (1.11) 4.19 (1.05)</td>
<td>3.94 (1.07) 4.09 (1.22)</td>
<td>1.03 –</td>
</tr>
<tr>
<td>Hastings is a leader rather than a follower.</td>
<td>Low WS 3.21 (1.07) 3.16 (1.25)</td>
<td>3.44 (1.941) 3.64 (1.17)</td>
<td>0.80 –</td>
</tr>
<tr>
<td>Hastings can be trusted.</td>
<td>Low WS 4.35 (0.984) 4.05 (0.916)</td>
<td>4.14 (0.941) 4.32 (1.38)</td>
<td>3.70 .055</td>
</tr>
<tr>
<td>Overall, I have a positive attitude toward Hastings.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If Hastings comes to my region I will buy from them.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* Mean (±std) of seven point scale anchored by 1 = “Strongly Disagree” and 7 = “Strongly Agree”.

*b* After mean centering by country by category.
“takes three wickets by three successive balls” which entitled the bowler to “be presented by his club with a new hat or some equivalent”. Subsequently the term spread to other sports, and today can refer to any “threefold feat”.

The reason a consumer, who knows little about mid-20th century art, can understand white space is the same reason why a football fan who knows little about mid-19th century cricket can understand a hat trick. That is, not knowing the origin of an expression (verbal or visual) does not impede one from understanding its current meaning. However, absent from that history, a hat trick would not mean what it does today. History matters even when the specifics are not known.

This simple but powerful insight into how visual language works seems to be less than universal in consumer research. Many still seem to believe that if a visual does not convey facts about a product through visual demonstrations or technical representations, then it simply elicits an effect or serves as a peripheral cue, often to the detriment of consumers’ understanding of the real message, i.e., the words. Even the “burning money” account of white space as narrow signal runs counter to such a limited view. We hope that by describing how consumers respond to one historically determined visual element, we will encourage more consumer researchers to think more about visuals as part of the living language of commercial communications. We also hope that this work encourages others to consider what they mean when the say or write “information.”

Acknowledgments

The authors would like to thank the Social Science and Humanities Resource Council of Canada (standard research grant #410-2010-1535) for supporting this research.

References


Patterns in consumption-based learning about brand quality for consumer packaged goods

Maciej Szymanowski a,⁎, Els Gijsbrechts b,1

a Rotterdam School of Management, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
b Quantitative Marketing at Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

ABSTRACT

In this paper, we explore the patterns of consumption-based learning about brand quality in mature consumer packaged goods (CPG) categories as well as category and household characteristics that drive such learning. We estimate brand-choice models with Bayesian learning on household purchases in over thirty CPG categories. This approach yields category- and household-specific estimates of the extent to which consumers update their knowledge on the quality of specific brands, with new consumption-based information from these brands. We then link this degree of learning to the underlying household and category drivers. We find that learning is present and significant in almost all categories yet varies in strength across categories and households. Learning about brand quality is negatively associated with variety seeking. Conversely, learning is stronger in categories where consumers have higher monetary (expensive items) and especially non-monetary stakes (categories with higher performance risk and involvement). In line with the ‘enrichment’ hypothesis, familiarity with the category resulting from frequent category purchases increases the information extracted from new consumption experiences — but only up to a certain point. Interestingly, however, market mavens learn less, potentially reflecting their overconfidence. Whereas some households learn more than others across the board, category factors are the strongest drivers of learning. Managerial implications are discussed. © 2013 Elsevier B.V. All rights reserved.

1. Introduction

Traditional consumer behavior models (Pechmann & Ratneshwar, 1992; Wright & Lynch, 1995) as well as studies in the field of information economics (Miller, 1984; Nelson, 1970) posit that consumer choices over time are driven by uncertainty and learning. Consumption-based learning about brands is a commonly acknowledged phenomenon in settings with ‘complex decision making’ (Narayanan, Chintagunta, & Miravete, 2007) and in high-involvement or strongly evolving categories such as insurance, pharmaceuticals, and telephone services (Akcura, Günl, & Petrova, 2004; Narayanan & Manchanda, 2009; Narayanan et al., 2007). However, there is growing evidence that even in mature consumer packaged goods (CPG) categories such as ketchup, liquid laundry detergent, toothpaste and toothbrushes, consumers may be uncertain about the product attribute levels that underlie the quality of a brand (Erdem, 1998; Erdem, Keane, & Sun, 2008; Mehta, Rajiv, & Srinivasan, 2004). This phenomenon is especially true for experience attributes, such as the teeth-whitening ability of a toothpaste brand or the cleansing power of a detergent, which may only reveal themselves over time or vary across usage occasions (Ching, Erdem, & Keane, 2012; Erdem, 1998). In the face of such uncertainty, consumers base their choices on beliefs about the brands’ quality 2 and rely on information obtained through usage to update these beliefs over time.

Not only is such consumption-based quality learning interesting academically, it is important from a managerial perspective. Every year, CPG manufacturers and retailers invest billions of dollars in enhancing the quality of their products and ensuring their consistency. Whether this money well spent depends on the extent to which consumers update their brand beliefs in response to quality improvements (Mitra & Golder, 2006) or breakdowns (van Heerde, Helsen, & Dekimpe, 2007). Moreover, experience-based brand-quality learning influences the impact of price promotions (Akcura et al., 2004; Erdem et al., 2008), free samples (Erdem, 1998), product innovation (Narayanan & Manchanda, 2009), and umbrella branding (Erdem, 1998; Erdem & Sun, 2002) and thus should play a role in the planning of these activities. Nonetheless, the available evidence on such learning in mature packaged goods is limited to only a few product categories, and little is known empirically about category or household characteristics that drive it.

2 Following extant literature on consumer brand choice under uncertainty, we interpret (perceived) quality as “a summary statistic that captures any intangible and tangible attributes of a product that may be imperfectly observable by consumers” (Erdem et al., 2004, p. 87).

3 For instance, in 2009 alone, Unilever invested 891 million Euros in R&D worldwide.
Our study aims to reduce this knowledge gap. In particular, we focus on the following questions. First, to what extent do consumers learn about the quality of specific brands from consumption experiences with those brands, in different CPG categories? Second, what are the drivers of such consumption-based learning? Specifically, which household and/or category characteristics determine the extent of brand-quality learning, and what is the direction and magnitude of this effect? The answers to these questions are important for both marketing scholars and practitioners. From an academic perspective, identifying categories where consumers learn the most about brand quality from consumption and assessing the category- or household-related antecedents enhances our understanding of this learning phenomenon. For managers, these insights may prove helpful in designing marketing programs. For instance, managers may focus learning-based tactics such as free samples for consumers on categories where consumption-based learning is stronger.

To address our research questions, we present a framework of the category- and household-factors that drive learning. This framework is empirically tested on data from a stable panel of 584 households, covering brand purchases in 33 product categories for a period of 3.5 years. We combine these purchase data with information from two surveys, documenting the panel member and category characteristics. We estimate choice models with Bayesian learning (see, e.g., Erdem & Keane, 1996; Narayanan & Manchanda, 2009) to obtain category- and household-specific learning parameters, which are then linked to their underlying factors. The results shed light on the prevalence and magnitude of learning for packaged goods, the variation in learning across categories and across households, and the underlying factors.

In the next section, we present our conceptual framework, building on the available literature on consumer decision making and learning. Section 3 describes the data and operationalization of the household and category drivers. The methodology is outlined in Section 4, followed by the results in Section 5. Section 6 summarizes the findings, discusses implications, and indicates limitations and suggestions for future research.

2. Conceptual framework

2.1. Background

Experience-based learning has long been a topic of interest in the consumer behavior and marketing modeling literature. Most of this literature has studied the early stages of learning in new or turbulent environments (Hutchinson & Eisenstein, 2008). In such settings, learning primarily involves the development of 'category knowledge': consumers have yet to uncover the relevant product attributes and their preferences regarding these attributes (Alba & Hutchinson, 1987). In this paper, we focus on mature consumer packaged goods (CPG) categories, for which such category knowledge is typically well developed. Instead, consumers primarily build 'brand knowledge': that is, they learn about the quality levels of specific brands (Kardes & Kalyanaram, 1992; Zhou, 2002). As in previous studies (e.g., Erdem et al., 2008; Narayanan et al., 2007), we assess such learning as the degree to which consumers update their beliefs about the quality of a brand based on consumption experiences with that brand, where 'quality' is some overall assessment of the experience attributes underlying the brand's 'performance' or 'appeal' (Erdem, 1998; Erdem & Keane, 1996).

Even in mature markets, consumers may not be perfectly knowledgeable about the quality of brands because the quality effects may take time or multiple consumptions to materialize (Erdem, 1998) or because it may be difficult to isolate the quality of the brand from other confounding factors (Hoch & Deighton, 1989). Furthermore, consumers generally buy and consume products sequentially rather than simultaneously, which hampers effective brand comparisons and learning (Hoch & Ha, 1986; Warlop, Ratneshwar, & van Osselaer, 2005) and causes the memory of brand quality to deteriorate between consumption and purchase occasions (Mehta et al., 2004; Warlop et al., 2005). Finally, consumers may need to update or revisit their knowledge as their consumption patterns evolve (Du & Kamakura, 2006).

Although these phenomena explain why consumers may gradually learn about brand quality from their experiences in CPG markets, they also suggest that the degree of learning may well vary across consumers and categories. Extant studies in consumer behavior and cognitive psychology conclude that, indeed, brand learning is both person- and product-dependent (see, e.g., Hoch & Deighton, 1989; Hoch & Ha, 1986; Hutchinson & Eisenstein, 2008; Warlop et al., 2005 for an overview). However, because their main focus is on uncovering (specific aspects of) the learning process as such, these papers are mostly conceptual in nature and/or use highly controlled laboratory settings to make their point. Conversely, analysis of scanner panel data has shown that consumption-based learning about brand quality is an empirically important phenomenon that may substantially affect brand choice over time (see, e.g., Ackerberg, 2003; Erdem & Keane, 1996; Mehta et al., 2004; Shin, Misra, & Horsky, 2012). Nonetheless, this evidence is confined to a limited number of categories and does not pursue household differences in learning. We aim to bridge these two literature streams by (i) empirically assessing consumers' experience-based brand-quality learning in a wide range of CPGs and (ii) testing a number of propositions on what drives the degree of learning.

2.2. Drivers of consumption-based learning

Building on the available literature on consumer decision making and learning from experience (e.g., Hoch & Deighton, 1989; Hutchinson & Eisenstein, 2008; Kardes, 1994), we postulate that the strength of learning about the quality of specific brands in a category depends on three dimensions: the consumers' domain familiarity, their incentives to learn, and their variety seeking. Below, we conceptualize how and why these dimensions trigger learning.

2.2.1. Domain familiarity

The degree to which consumers learn from new consumption signals of specific brands depends on their 'domain familiarity' — that is, their accumulated experiences with the category (Alba & Hutchinson, 1987; Hoch & Deighton, 1989; Hutchinson & Eisenstein, 2008). The general premise — also known as the 'enrichment hypothesis' (Johnson & Russo, 1984) — is that category familiarity creates 'expertise,' which in turn allows consumers to better comprehend new consumption signals and 'retain more information with lower levels of error' (Hutchinson & Eisenstein, 2008, p. 103). For one, consumers who are more familiar with the product category have acquired more 'cognitive structure' and thus are better able to organize information from new experiences (Hutchinson & Eisenstein, 2008). Moreover, greater knowledge of relevant product attributes and better developed preferences for these attributes greatly reduce their cognitive load and the (mental) cost of processing new signals (Bronnitzer, 2008). Experts have also been found to engage in more 'analytic processing,' i.e., to be more influenced by brand-specific attributes rather than by salient but irrelevant overall impressions (Hutchinson & Alba, 1991; Hutchinson & Eisenstein, 2008).

4 Examples of such studies are Carpenter and Nakamoto (1989, 1996), who study how experiences with pioneer brands shape consumers' preferences in a new product class, or Heilman, Bowman, and Wright (2000), who analyze how category experience drives choice dynamics for consumers that are new to a market.

5 Moreover, the results of these studies may be difficult to compare because of (i) differences in the definition of consumption units and categories, (ii) variations in country/type of brand analyzed, and (iii) methodological differences (i.e., inclusion of forgetting and consumer heterogeneity).
This finding suggests that consumers who are more familiar with the product category extract more information from a given consumption signal. However, there are also counter-indications. Experienced buyers may be less likely to encode new signals that make earlier information obsolete (Wood & Lynch, 2002), and their 'overconfidence' might slow the dissemination of new information (Hoch & Deighton, 1989; Hutchinson & Eisenstein, 2008). Such a finding would imply that consumers highly familiar with the category learn less about brand quality from consumption than moderately familiar consumers (Johnson & Russo, 1984). Given these countervailing forces, we postulate that product category familiarity affects the informativeness of consumption signals but leave the direction of the effect an empirical issue.

2.2.2. Incentives to learn

We expect consumer learning to be positively related to the anticipated gains from learning. This claim stems from the notion in information economics that learning comes at a cost, which is justified if consumers expect to obtain sufficient gains (Stigler, 1961). Greater gains from learning may incite consumers to pay more attention to the actual performance of a brand (i.e., intentional learning increases the likelihood of analytic information processing, Hutchinson & Alba, 1991) and thus learn more (Hawkins & Hoch, 1992). Incentives to learn may stem from monetary as well as non-monetary gains. First, we expect consumers to have higher learning stakes in 'expensive' categories that require a high discretionary sum spent per purchase occasion (Narasimhan, Neslin, & Sen, 1996). Second, incentives to learn may be related to the performance risk or the expected utility loss in case of the wrong brand choice in the category (Beatty & Smith, 1987). More performance risk in a category will lead stronger households to be more motivated to learn from experience to avoid mistakes. Finally, consumers will be more keen to monitor brand quality experiences and more likely to use accessible content in categories in which they are more involved (Celsi & Olson, 1988; Schwarz, 2004).

2.2.3. Brand loyalty/variety seeking

The degree of brand loyalty or, conversely, variety seeking is the third dimension we associate with the magnitude of brand-quality learning. Whereas consumption-based learning is driven by the desire to maximize the expected quality of brands consumed, variety seeking is associated with the desire to mitigate boredom and satiation (Gijssbrechts, Campo, & Nisol, 2000; McAlistier & Pessemier, 1982). Previous research has documented that consumers with an intrinsic desire for stimulation may be willing to give up quality in return for variety (Kahn & Raju, 1991; Trivedi & Morgan, 2003). Therefore, we expect such variety-seeking consumers to take less interest in learning about the quality of specific brands.

3. Data and variable operationalizations

3.1. Data

3.1.1. Scanner panel data

Our main data source consists of GfK scanner panel data covering households' supermarket purchases, over 3.5 years, across a wide range of product categories. Only households that are active in the panel throughout the entire period are retained. We use the first year of data as an initialization period, maintaining the remaining 2.5 years for model estimation. Given our interest in purchase dynamics, households are only included in a category if they have at least 6 purchases in that category in the estimation sample. The final estimation sample thus covers 584 different households engaging in 154,077 purchases across 33 categories, as defined by GfK.

Our data include brand-level information on the quantities bought and the corresponding prices and promotions. In each category, we consider the top brands covering at least 50% of total category sales (with a minimum of four brands) for inclusion in the brand choice model. Table 1 provides descriptions for the brands in our sample categories. This table indicates that categories vary considerably in terms of concentration (the share of sampled purchases of the top 2 brands ranging from 41% in the powdered laundry detergent category to 85% for sugar) and price spread (the highest-to-lowest price ratio, ranging from 1.06 for dairy products to 3.86 for flavor additives). In addition, the data show considerable variation in households' brand choices over time—a desirable feature for the assessment of learning.6

3.1.2. Survey data

In addition to the scanner panel data, we have access to information from two surveys, which we use to operationalize household and category drivers. The first survey, administered to the panel members by GfK, records households' time-invariant scores on a number of attitudinal and shopping-related items. The second survey, administered7 to Dutch consumers who are not members of the panel, yields summary measures on a number of category characteristics that do not vary across households or over time. Below, we indicate how these data are used to operationalize the drivers of learning.

3.2. Driver operationalizations

Table 2 provides an overview of the learning drivers, together with the underlying dimensions and the hypothesized effects. For each driver, the table also provides measurement details and indicates the corresponding data source. We briefly discuss the learning drivers below.

3.2.1. Domain familiarity

We measure a household’s ‘domain familiarity’ or accumulated category knowledge (CatKnow6) by the average number of category purchases per month in the year preceding our estimation sample. As indicated above, this driver’s impact on learning may be positive (e.g., better comprehension of new brand-specific signals) or negative (e.g., less encoding of such signals), and the direction of the effect may depend on the knowledge level already attained. To flexibly capture its impact on learning, we include both the variable and its square in the model.

3.2.2. Incentives to learn

Category expensiveness (Expensiveness6) is measured as the average amount spent by the household per purchase occasion in the category (Narasimhan et al., 1996). Performance risk in a given category (Risk,6) and category involvement (Involvement,6) are captured through category-specific item scores from the second survey (see Table 2 for details). Based on our conceptualization, we expect all three variables to be positively associated with learning.

3.2.3. Variety seeking

Building on the available literature, we adopt three indicators of brand loyalty/variety seeking. The first, Brand_Loyalty6, is a household trait obtained from the panel-member survey data (see Table 2 for details). This indicator reflects the household’s commitment to his/her favorite brand in different categories (see, e.g., Baumgartner & Steenkamp, 1996) — with an expected positive impact on learning. The second indicator, Need_Variety,6 measures the need for variety as a category characteristic (Givon, 1984; Pessemier & Handelsman, 1984). We expect higher values for this variable to entail less learning. Third, we include the number of different brands purchased by a household in a given category (Number_Brands6), which we obtain for the year prior to our estimation sample and mean-center against the category average. This measure is behavioral in nature (van Trijp & Steenkamp, 2004).

6 Please see Appendix B for details on model identification.
7 These data were collected in the context of a study by Steenkamp et al. (2004). We thank the authors for making part of the information available to us for this research.
Table 1
Descriptive statistics for brands in sample categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Brand</th>
<th>Choice share</th>
<th>Temporal variation in households' choice shares</th>
<th>Number of purchases</th>
<th>Average price (in Eurocent)</th>
<th>Brand type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavor additives</td>
<td>1</td>
<td>0.08</td>
<td>0.06</td>
<td>194</td>
<td>0.41/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.20</td>
<td>0.12</td>
<td>483</td>
<td>1.43/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15</td>
<td>0.10</td>
<td>350</td>
<td>0.46/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.10</td>
<td>0.07</td>
<td>249</td>
<td>0.37/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.15</td>
<td>0.10</td>
<td>365</td>
<td>0.79/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.31</td>
<td>0.19</td>
<td>733</td>
<td>0.38/g</td>
<td>NB</td>
</tr>
<tr>
<td>Periodic care/diapers</td>
<td>1</td>
<td>0.13</td>
<td>0.05</td>
<td>490</td>
<td>21.96/ piece</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.13</td>
<td>0.10</td>
<td>506</td>
<td>20.15/ piece</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.31</td>
<td>0.16</td>
<td>1188</td>
<td>20.05/ piece</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.42</td>
<td>0.15</td>
<td>1594</td>
<td>21.03/ piece</td>
<td>NB</td>
</tr>
<tr>
<td>Dairy products</td>
<td>1</td>
<td>0.16</td>
<td>0.04</td>
<td>1521</td>
<td>0.43/ml</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.33</td>
<td>0.10</td>
<td>3247</td>
<td>0.44/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.24</td>
<td>0.08</td>
<td>2333</td>
<td>0.46/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.27</td>
<td>0.09</td>
<td>2619</td>
<td>0.45/ml</td>
<td>NB</td>
</tr>
<tr>
<td>Air refreshers</td>
<td>1</td>
<td>0.29</td>
<td>0.16</td>
<td>733</td>
<td>2.08/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.50</td>
<td>0.19</td>
<td>1244</td>
<td>2.85/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.13</td>
<td>0.12</td>
<td>314</td>
<td>2.84/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.08</td>
<td>0.04</td>
<td>194</td>
<td>1.42/g</td>
<td>PL</td>
</tr>
<tr>
<td>Dairy drinks</td>
<td>1</td>
<td>0.13</td>
<td>0.05</td>
<td>1711</td>
<td>0.11/ml</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.32</td>
<td>0.09</td>
<td>3087</td>
<td>0.12/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15</td>
<td>0.07</td>
<td>1479</td>
<td>0.13/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.19</td>
<td>0.04</td>
<td>1800</td>
<td>0.11/ml</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.16</td>
<td>0.04</td>
<td>1538</td>
<td>0.1/ml</td>
<td>PL</td>
</tr>
<tr>
<td>Liquid laundry detergent</td>
<td>1</td>
<td>0.39</td>
<td>0.18</td>
<td>1132</td>
<td>0.71/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.16</td>
<td>0.13</td>
<td>447</td>
<td>0.69/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15</td>
<td>0.05</td>
<td>433</td>
<td>0.44/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.14</td>
<td>0.13</td>
<td>394</td>
<td>0.85/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.17</td>
<td>0.12</td>
<td>477</td>
<td>0.69/ml</td>
<td>NB</td>
</tr>
<tr>
<td>Salads</td>
<td>1</td>
<td>0.19</td>
<td>0.10</td>
<td>1146</td>
<td>0.82/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.21</td>
<td>0.13</td>
<td>1267</td>
<td>0.86/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.11</td>
<td>720</td>
<td>0.63/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.48</td>
<td>0.20</td>
<td>2903</td>
<td>0.81/g</td>
<td>NB</td>
</tr>
<tr>
<td>Meal decorators</td>
<td>1</td>
<td>0.33</td>
<td>0.14</td>
<td>2714</td>
<td>2.08/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.39</td>
<td>0.13</td>
<td>3272</td>
<td>1.75/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.14</td>
<td>0.08</td>
<td>1124</td>
<td>2.05/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.14</td>
<td>0.10</td>
<td>1190</td>
<td>1.94/g</td>
<td>NB</td>
</tr>
<tr>
<td>Breakfast cereals</td>
<td>1</td>
<td>0.26</td>
<td>0.12</td>
<td>1120</td>
<td>0.83/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.20</td>
<td>0.11</td>
<td>880</td>
<td>0.75/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.05</td>
<td>526</td>
<td>0.58/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.31</td>
<td>0.08</td>
<td>1334</td>
<td>0.64/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.11</td>
<td>0.05</td>
<td>487</td>
<td>0.62/g</td>
<td>NB</td>
</tr>
<tr>
<td>Mouth hygiene</td>
<td>1</td>
<td>0.50</td>
<td>0.14</td>
<td>1374</td>
<td>2.28/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.14</td>
<td>0.08</td>
<td>390</td>
<td>4.55/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.22</td>
<td>0.10</td>
<td>611</td>
<td>2.58/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.14</td>
<td>0.08</td>
<td>373</td>
<td>3.3/ml</td>
<td>NB</td>
</tr>
<tr>
<td>Pet food</td>
<td>1</td>
<td>0.09</td>
<td>0.04</td>
<td>690</td>
<td>0.44/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.08</td>
<td>0.05</td>
<td>612</td>
<td>0.49/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.23</td>
<td>0.09</td>
<td>1747</td>
<td>0.47/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.14</td>
<td>0.05</td>
<td>1079</td>
<td>0.69/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.28</td>
<td>0.11</td>
<td>2060</td>
<td>0.52/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.17</td>
<td>0.09</td>
<td>1274</td>
<td>0.38/g</td>
<td>NB</td>
</tr>
<tr>
<td>Fabric softeners</td>
<td>1</td>
<td>0.09</td>
<td>0.06</td>
<td>311</td>
<td>0.31/ml</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.44</td>
<td>0.20</td>
<td>1447</td>
<td>0.36/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.19</td>
<td>0.07</td>
<td>630</td>
<td>0.2/ml</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.28</td>
<td>0.17</td>
<td>907</td>
<td>0.31/ml</td>
<td>NB</td>
</tr>
<tr>
<td>Spices and herbs</td>
<td>1</td>
<td>0.14</td>
<td>0.08</td>
<td>610</td>
<td>6.06/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.63</td>
<td>0.18</td>
<td>2714</td>
<td>8.16/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.06</td>
<td>0.07</td>
<td>267</td>
<td>3.89/g</td>
<td>NB</td>
</tr>
<tr>
<td>Alcoholic drinks</td>
<td>1</td>
<td>0.54</td>
<td>0.15</td>
<td>3413</td>
<td>0.61/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.16</td>
<td>0.09</td>
<td>1004</td>
<td>0.23/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.08</td>
<td>751</td>
<td>0.24/ml</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.18</td>
<td>0.09</td>
<td>1098</td>
<td>0.24/ml</td>
<td>NB</td>
</tr>
<tr>
<td>Dish detergent (tablets)</td>
<td>1</td>
<td>0.19</td>
<td>0.04</td>
<td>259</td>
<td>1.45/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.09</td>
<td>0.04</td>
<td>132</td>
<td>0.71/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.48</td>
<td>0.18</td>
<td>670</td>
<td>1.14/g</td>
<td>NB</td>
</tr>
<tr>
<td>Powdered laundry detergent</td>
<td>1</td>
<td>0.11</td>
<td>0.07</td>
<td>297</td>
<td>1.07/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.19</td>
<td>0.15</td>
<td>497</td>
<td>0.43/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.21</td>
<td>0.13</td>
<td>557</td>
<td>0.43/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.17</td>
<td>0.05</td>
<td>432</td>
<td>0.26/g</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.12</td>
<td>0.08</td>
<td>313</td>
<td>0.59/g</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.20</td>
<td>0.16</td>
<td>514</td>
<td>0.38/g</td>
<td>NB</td>
</tr>
<tr>
<td>Margarine/oil</td>
<td>1</td>
<td>0.09</td>
<td>0.05</td>
<td>811</td>
<td>0.51/g</td>
<td>PL</td>
</tr>
<tr>
<td>Category</td>
<td>Brand</td>
<td>Choice share</td>
<td>Temporal variation in households' choice shares*</td>
<td>Number of purchases</td>
<td>Average price (in Eurocent)</td>
<td>Brand type</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------</td>
<td>--------------</td>
<td>-----------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Shares in households</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other than PL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and reflects the household’s purchase variety in the category relative to other households. We expect higher levels of this variable to be associated with lower levels of learning.

3.2.4. Control variables

Apart from these hypothesized drivers, we include several control variables. First, we distinguish between food and non-food categories (Foodc equals 1 for food categories and zero otherwise). Next, we include two household characteristics. Mavenismh reflects the extent to which the consumer acts as an opinion leader and shares his/her product information and experience with other consumers (Feick & Price, 1987) and is obtained as a summated scale from the panel members’ self-reported items in the first survey (see Table 2). Because we expect shopping mavens to be more involved grocery shoppers (who are more eager to learn) but also more knowledgeable shoppers (who have little left to learn), we offer no directional hypotheses on their degree of consumption-based learning. Planningh captures the household’s degree of purchase planning through the use of a mental or written shopping list (Inman, Winer, & Ferraro, 2009) and is also obtained from the panel-member survey (see Table 2). Buyers who plan their specific purchases in advance are likely to be more certain about their choices and, as indicated by Urbany (1986) and Urbany, Dickson, and Wilkie (1989), tend to be less open to new information and less responsive to changes in utility. This characteristic would imply a negative relationship between planning and subsequent, experience-based learning. However, such a link only holds to the extent that consumers plan the specific brand within the category. Because our measure does not contain this level of detail (i.e., it pertains to the consumers’ degree of planning across product categories and shopping trips), it enters as a control only.

### Table 2

Overview of learning drivers.

<table>
<thead>
<tr>
<th>Underlying dimension</th>
<th>Variable: Notation</th>
<th>Expected effect on learning</th>
<th>Data source</th>
<th>Source of variation</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain familiarity</td>
<td>Category purchase frequency: CatKnowc</td>
<td>+/−</td>
<td>Scanner data</td>
<td>Households and categories</td>
<td>Household h’s average number of purchases/month in category c, in a one-year initialization period.</td>
</tr>
<tr>
<td>Incentives to learn</td>
<td>Category expensiveness: Expensivenessc</td>
<td>+</td>
<td>Scanner data</td>
<td>Households and categories</td>
<td>Average amount spent by household h per purchase occasion in category c, in a one-year initialization period.</td>
</tr>
<tr>
<td>Incentives to learn</td>
<td>Performance risk: Riskc</td>
<td>+</td>
<td>Non-panel-member survey</td>
<td>Categories</td>
<td>Score average of the following items for category c:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- There is much to lose if you make the wrong choice in category c</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- It matters a lot when you make the wrong choice in category c</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- In category c, there are big differences in quality between the various brands</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Scale Average obtained from Steenkamp, Geyskens, Gielen, &amp; Koll, 2004)</td>
</tr>
<tr>
<td>Incentives to learn</td>
<td>Category involvement: Involvem,</td>
<td>+</td>
<td>Non-panel-member survey</td>
<td>Categories</td>
<td>Score average of the following items for category c:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- The category c is very important to me</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- The category c interests me a lot</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Scale Average obtained from Steenkamp et al., 2004)</td>
</tr>
<tr>
<td>Variety seeking</td>
<td>Number of brands consumed: Number_Brandsc</td>
<td>−</td>
<td>Scanner data</td>
<td>Households and categories</td>
<td>Number of different brands bought by household h in category c, in a one-year initialization period (mean-centered relative to the category mean across households).</td>
</tr>
<tr>
<td>Variety seeking</td>
<td>Brand loyalty: Brand_Loyaltyh</td>
<td>+</td>
<td>Panel-member survey</td>
<td>Households</td>
<td>Score average for household h on the following items:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I prefer to buy the brand that I usually buy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I consider myself a brand-loyal consumer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I really feel connected with brands that I buy (Cronbach’s Alpha = .953)</td>
</tr>
<tr>
<td>Variety seeking</td>
<td>Need for variety: Need Variety,</td>
<td>−</td>
<td>Non-panel-member survey</td>
<td>Categories</td>
<td>Score of the following item for category c:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In category c, I want a good variety of product variants to choose from Dummy equal to 1 if c is a food category and to 0 if c is a non-food category</td>
</tr>
<tr>
<td>Control variable</td>
<td>Dummy for food categories: Foodc</td>
<td>+/−</td>
<td>Scanner data</td>
<td>Categories</td>
<td>Score average for household h on the following items:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I talk with friends about products I buy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Friends and neighbors often ask me for advice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- People often ask me what I think about new products (Cronbach’s Alpha = .969)</td>
</tr>
<tr>
<td>Control variable</td>
<td>Mavenismh: Mavenismh</td>
<td>+/−</td>
<td>Panel-member survey</td>
<td>Households</td>
<td>Score average for household h on the following items:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- When I go shopping I know exactly what I want to buy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I plan my purchases carefully so that supermarket visits will cost less time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- I make a list before I go shopping and stick to it (Cronbach’s Alpha = .953)</td>
</tr>
</tbody>
</table>

### Table 1 (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Brand</th>
<th>Choice share</th>
<th>Temporal variation in households’ choice shares</th>
<th>Number of purchases</th>
<th>Average price (in Eurocent)</th>
<th>Brand type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deodorant</td>
<td>1</td>
<td>0.18</td>
<td>0.07</td>
<td>232</td>
<td>95.30/piece</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.23</td>
<td>0.12</td>
<td>293</td>
<td>90.28/piece</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.21</td>
<td>0.07</td>
<td>268</td>
<td>83.40/piece</td>
<td>PL</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.17</td>
<td>0.08</td>
<td>218</td>
<td>99.18/piece</td>
<td>NB</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.21</td>
<td>0.14</td>
<td>276</td>
<td>176.50/piece</td>
<td>NB</td>
</tr>
</tbody>
</table>

* a Standard deviation of the household’s brand share, calculated over four 32-week subperiods, and then averaged over households visiting the chain.

* b Based on a ten-point scale (1 = strongly disagree to 10 = strongly agree).

* c Based on a five-point scale (1 = strongly disagree to 5 = strongly agree).
4. Methodology

4.1. Stage 1: Degree of consumption-based quality learning

In the first stage, we estimate the extent to which households learn about brand quality from consumption signals in different categories.

4.1.1. Basic model specification

Our approach to quantifying the amount of consumption-based quality learning closely follows the available literature and is based on structural dynamic brand choice models with Bayesian learning (see, e.g., Crawford & Shum, 2005; Erdem & Keane, 1996; Mehta et al., 2004; Narayanan & Manchanda, 2009). In these models, consumers select the brand from a category that maximizes their current utility, the utility for household \( h \) of brand \( j \) in category \( c \) on purchase occasion \( t \) being given by the following:

\[
U^h_{jct} = f \left( Q^h_{jct} \right) + X^h_{jct} \beta^h_c + \epsilon^h_{jct}.
\]

where \( Q^h_{jct} \) indicates household \( h \)'s beliefs on purchase occasion \( t \) regarding brand \( j \)'s 'quality', i.e., a composite of the brand's experience attributes about which the household is uncertain.\(^8\) \( X^h_{jct} \), in turn, is a vector of utility determinants other than perceived quality that are directly observable to both the researcher and the consumer. Similar to previous brand choice models, these determinants include price and promotion support (feature and display).\(^9\)

To accommodate the fact that the attributes of a brand may be better tuned to the needs of some consumers than others,\(^10\)

Information about these attributes is accumulated through actual brand experience. Specifically, consumption by household \( h \) of each unit of brand \( j \) in category \( c \) in period \( t - 1 \) provides a new quality signal \( g^h_{jct} \), which is assumed to be i.i.d. normally distributed with a mean equal to the true brand quality \( q^h_j \) and variance \( \lambda^h_{jt} \), or:

\[ g^h_{jct} \sim N \left( q^h_j, \lambda^h_{jt} \right). \]

We denote a series of \( M_h \) consumption units for household \( h \) of brand \( j \) in category \( c \) consumed at time \( t - 1 \) as

\[
G^h_{jct} = \sum_{t' = m-1}^{t-1} g^h_{jct} - N \left( q^h_j, \lambda^h_{jt} \right). \]

Consumers' uncertainty is reduced gradually as they learn, but it may also increase again due to forgetting (Mehta et al., 2004).\(^11\) In the absence of consumption at \( t - 1 \), we expect \( \lambda^h_{jt} \) to increase exponentially:

\[ \lambda^h_{jt} = \lambda^h_{jt-1} \times e^{\beta^h_t (w^h_t - \lambda^h_{jt-1})}, \]

where \( \beta^h_t \) is an estimated decay parameter for household \( h \) and category \( c \) and \( w^h_t - \lambda^h_{jt-1} \) refers to the time elapsed between purchase occasions \( t \) and \( t - 1 \).\(^12\) As in previous studies, we assume that on each occasion \( t \) in category \( c \), the consumer adopts only one brand, such that \( \sum_j d^h_{jct} = 1 \), where \( d^h_{jct} = 1 \) if brand \( j \) in category \( \sim \)


\[ \text{in the consumer's preferences. However, to the extent that consumers' preferences in the consumer's uncertainty is reduced gradually as they learn, but it may also increase again due to forgetting (Mehta et al., 2004).} \]
Table 3
Notation overview for the first stage model.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>Index for product category</td>
</tr>
<tr>
<td>c</td>
<td>Number of sample categories</td>
</tr>
<tr>
<td>j, k</td>
<td>Indices for brand</td>
</tr>
<tr>
<td>t, h</td>
<td>Indices for purchase occasion and household</td>
</tr>
<tr>
<td>d_{ij}</td>
<td>Indicator equal to 1 if brand j in category c was consumed by household h at purchase occasion t and 0 otherwise</td>
</tr>
<tr>
<td>D_{ij}</td>
<td>Vector of purchase indicators d_{ij} up to purchase occasion t</td>
</tr>
<tr>
<td>E_{ij}</td>
<td>Vector of consumption signals C_{ij} up to purchase occasion t</td>
</tr>
<tr>
<td>J_{ij}</td>
<td>Number of sample brands, and number of households, respectively, in category c</td>
</tr>
<tr>
<td>m</td>
<td>Index for unit of a product</td>
</tr>
<tr>
<td>M_{ij}</td>
<td>Number of product units purchased by household h in category c on the purchase occasion t</td>
</tr>
<tr>
<td>T_{ij}</td>
<td>Number of purchases in category c made by household h during the sample period</td>
</tr>
<tr>
<td>w_{j}</td>
<td>Week index for purchase t of household h in category c</td>
</tr>
<tr>
<td>Z_{ij}</td>
<td>Indicator equal to 1 if brand j in category c was available to household h at purchase occasion t and 0 otherwise</td>
</tr>
</tbody>
</table>

Utility and its determinants

- \( \lambda_{ij} \): Risk aversion parameter of household h in category c
- \( U_{ij} \): Utility of household h for brand j in category c at purchase occasion t
- \( X_{ij} \): Utility determinants for brand j in category c at purchase occasion t other than quality belief
- \( \beta_{ij} \): Household's sensitivity parameters to utility determinants in X_{ij}
- \( \epsilon_{ij} \): i.i.d. extreme value component of utility observed by the household h, unobserved by the researcher

Quality and quality belief

- \( \xi_{ij} \): True quality of brand j in category c for household h
- \( Q_{ij} \): Quality belief of household h for brand j in category c on purchase occasion t
- \( \mu_{ij} \): Mean quality belief of household h for brand j in category c on purchase occasion t
- \( \alpha_{ij} \): Uncertainty or variance in quality belief of household h for brand j in category c at purchase occasion t (Initial belief variance: \( \alpha_{ij0}^2 \))

Learning and forgetting

- \( b_{ij} \): Parameter capturing the rate of forgetting for household h in category c
- \( t_{ij} \): Household h's information set in category c at purchase occasion t
- \( g_{ijn} \): Consumption signal of one unit of brand j in category c on purchase occasion t by household h
- \( G_{ijt} \): Average over a series of M_{ijt}'s consumption signals g_{ijn}, for brand j in category c on purchase occasion t
- \( \lambda_{ij}^2 \): Standard deviation of the consumption signal in category c for household h

Household heterogeneity

- \( \nu_{ij}, \varsigma_{ij} \): Mean and variance across households of the parameter capturing the extent of forgetting, i.e., \( b_{ij}^2 = \log -N(\nu_{ij}, \varsigma_{ij}) \)
- \( \nu_{ij}, \varsigma_{ij} \): Mean and variance across households of the true quality parameter, i.e., \( \xi_{ij}^2 = \log -N(\nu_{ij}, \varsigma_{ij}) \)
- \( \nu_{ij}, \varsigma_{ij} \): Mean and variance across households of the risk aversion parameter, i.e., \( \beta_{ij}^2 = \log -N(\nu_{ij}, \varsigma_{ij}) \)
- \( \nu_{ij}, \varsigma_{ij} \): Vectors of mean and variance across households of the parameters \( b_{ij}^2 \), \( \xi_{ij}^2 \), \( \beta_{ij}^2 \), \( \nu_{ij}, \varsigma_{ij} \)
- \( \nu_{ij}, \varsigma_{ij} \): Mean and variance across households of the standard error of the consumption signal, i.e., \( \lambda_{ij}^2 \)
- \( \nu_{ij}, \varsigma_{ij} \): Set of choice model parameters to be estimated (including the means and variances of the mixing distributions):

\[ \alpha_{ij} = \{ \nu_{ij}, \varsigma_{ij} \} \]

Effects specification. The parameters \( g_{ijn} \) and \( \lambda_{ij} \) are normally distributed, whereas \( \lambda_{ij}^2, b_{ij}^2, \) and \( \xi_{ij}^2 \) have log-normal distributions (to ensure positive values). The initial quality belief is assumed to be normally distributed, with mean \( \mu_{ij0} \) and variance \( \sigma_{ij0}^2 \). We assume that at time \( t = 0 \), consumers are uncertain about their own 'true quality' but know the brand's mean true quality across the population (\( \nu_{ij0} \)) and rationally form their initial mean beliefs to be the same as this population-level mean, such that \( \mu_{ij0} = \nu_{ij0} \). (see Crawford & Shum, 2005; Narayanan & Manchanda, 2009 for a similar approach). The initial belief uncertainty is estimated as a separate parameter \( \sigma_{ij0} \), which, for reasons of parsimony, we set to be homogeneous across brands (see also Erdem & Keane, 1996; Mehta et al., 2004). We estimate the parameters using simulated likelihood. Appendix A provides details pertaining to the log likelihood function and the estimation procedure.

4.1.2. Degree of learning

As shown in Eq. (4), when the variance of the consumption signal \( \lambda_{ij}^2 \) is smaller, the impact of a new consumption signal \( g_{ijn} \) on the consumer's quality belief \( \xi_{ij} \) increases, and vice versa. Therefore, a smaller (larger) consumption signal variance implies stronger (weaker) learning from consumption. However, Eq. (4) also reveals that the degree of updating depends on the prior brand quality uncertainty, which amounts to \( \sigma_{ij0} \) at the beginning of the observation period. Because this initial variance also differs across categories, we 'normalize' the consumption signal variance against this initial uncertainty. Therefore, the estimated value of the consumption signal variance relative to the initial variance (\( \lambda_{ij}^2/\sigma_{ij0}^2 \)) captures the magnitude of learning and thus constitutes the focal parameter in our study. For convenience, we refer to this ratio as the 'learning parameter', with higher levels of this parameter implying less learning.

The learning parameters thus reflect the informativeness of one consumption signal,14 where each consumption signal in a category represents the experience contained in one consumption unit in that category. We normalize the measurement units such that a category's 'consumption unit' corresponds to the average purchase quantity, across consumers, per purchase occasion in that category. More specifically, we specify the number of normalized units adopted by the household in category c at time t as \( M_{ijt} = \nu_{ij0} \cdot \Gamma_{ij} \), where \( M_{ij0} \) is the quantity adopted by the household at that time and \( \Gamma_{ij} \) is the average category purchase quantity across households and purchase occasions (which largely coincides with a 'standard' or most commonly adopted package size). For instance, if household h bought 500 g of coffee at time t (\( M_{ij0} = 500 \) g) and the average purchase quantity equals 250 g (\( \Gamma_{ij} = 250 \) g), then the normalized number of units for the household at that time becomes 2 (\( M_{ijt} = 2 \)). This normalization will enable meaningful cross-category comparisons.

4.2. Stage 2: Linking consumption-based learning to household and category drivers

In stage 2, we examine how household and category drivers influence the degree of learning. The dependent variable in this stage is the opposite of the logarithm of the learning parameter, obtained from the first-stage equation. Specifically, for each household and category, we calculate the posterior estimate of the (logarithm of the) learning parameter as a weighted average of draws15 from the parameter mixing distribution in that category, using the household-specific likelihoods as weights (see Train, 2003, p. 266). To facilitate interpretation of the coefficients, we use the negative of the log-learning parameter, \(-\log(\lambda_{ij}^2/\sigma_{ij0}^2)\), for our dependent variable, such that positive coefficients associated with a driver point to stronger learning. As explanatory variables, we use the household characteristics ('Brand_Loyalty', 'Mavenism', 'Planning'), category characteristics ('Risk', 'Need Variety', 'Food', 'Involvement'), and household- as well as category-specific

14 Mehta, Rajiv, and Srinivasan (2003) refer to it as the inverse of the 'signal-to-noise ratio', which measures the extent to which the initial uncertainty of a consumer differs from the "inherent product variability."

15 Specifically, for each category and household, we consider multiple draws from the mixing distribution of the consumption signal variance. Each of these draws is then divided by the (homogeneously) estimated initial belief variance and log-transformed to obtain the log-learning parameter. We then 'weight' this log-learning parameter with the household's likelihood function evaluated at the corresponding consumption signal variance draw and initial variance estimate before aggregating across draws.
variables (CatKnow, Expensiveness, Number_Brands) from our conceptual framework. This approach leads to the following equation:

\[ -\log(\lambda^2/\sigma^2_{\theta}) = \theta_0 + \theta_1 \times Brand\_Loyalty + \theta_2 \times Mavenism + \theta_3 \times Planning + \theta_4 \times Risk + \theta_5 \times Need\_Variety + \theta_6 \times Food + \theta_i \times Involvement + \theta_8 \times CatKnow + \theta_9 \times Expensiveness + \theta_{10} \times Number\_Brands + \omega^2, \]

(6)

where \(\log(\lambda^2/\sigma^2_{\theta})\) is the posterior log-learning parameter for household \(h\) and category \(c\) (obtained from stage 1), \(\theta\) are parameters capturing its sensitivity to the household-, category-, or household- and category-specific drivers, and \(\omega\) is a household- and category-specific error.

To ensure the efficiency of the estimates, we must consider that our dependent variable is a posterior of an estimated parameter, which is inherently uncertain. To assess the sampling error, we compute for each category and household a set of posterior values of the learning parameter using draws from model estimates’ sampling distributions. We then compute, for each category and household, the standard deviation of the resulting posteriors of the learning parameter and use their inverse as weights in the second-stage regression (see Steenkamp, Nijss, Hanssens, & Dekimpe, 2005 for a similar approach).

5. Results

5.1. Stage 1: Degree of consumption-based quality learning

5.1.1. Model fit

In each category, we compare the fit of the proposed Bayesian learning model to a benchmark model with no purchase dynamics (Appendix B reports the fit statistics). Except in one category (meal decorators), the model with Bayesian quality updating results in an average of household and category combinations, learning is substantial.

5.1.2. Learning parameter estimates

Table 4 lists our sample categories, together with their estimated learning parameters, sorted from lowest to highest (the complete set of stage-1 estimation results, for all categories, is given in Appendix C).

As the parameter follows an asymmetric (log-normal) distribution across households, we focus on the median rather than the mean as a summary statistic. For each category, we also report the lower and upper quartile of the learning parameter to provide a feel for the differences between households. The mean value of the learning parameter, pooled across categories and households, is .62. Given that this figure reflects the variance in the consumption signals relative to the initial uncertainty, its low value implies that new information obtained through consumption is highly influential. Our observed rate of learning appears within the ‘ballpark’ range from earlier studies (see Erdem, 1998; Erdem & Sun, 2002; Shin et al., 2012) and exerts a strong influence on the updated beliefs. Specifically, the consumer’s first consumption experience determines 62% of his updated quality belief (the prior belief determining the remaining 38%)—be it that this figure levels off with subsequent brand purchases. This finding suggests that for the average of household and category combinations, learning is substantial.

At the same time, Table 4 points to considerable variation across categories and across households within categories. Categories with strong learning (low learning parameter) appear to be mostly

Table 4

<table>
<thead>
<tr>
<th>Category</th>
<th>Learning parameter: distribution within category across households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
</tr>
<tr>
<td>Flavors</td>
<td>.000</td>
</tr>
<tr>
<td>Periodic care/diapers</td>
<td>.001</td>
</tr>
<tr>
<td>Dairy products</td>
<td>.001</td>
</tr>
<tr>
<td>Air refreshers</td>
<td>.002</td>
</tr>
<tr>
<td>Dairy drinks</td>
<td>.008</td>
</tr>
<tr>
<td>Liquid laundry detergent</td>
<td>.019</td>
</tr>
<tr>
<td>Salads</td>
<td>.019</td>
</tr>
<tr>
<td>Meal decorators</td>
<td>.020</td>
</tr>
<tr>
<td>Breakfast cereals</td>
<td>.022</td>
</tr>
<tr>
<td>Mouth hygiene</td>
<td>.025</td>
</tr>
<tr>
<td>Pet food</td>
<td>.026</td>
</tr>
<tr>
<td>Fabric softeners</td>
<td>.044</td>
</tr>
<tr>
<td>Spices and herbs</td>
<td>.049</td>
</tr>
<tr>
<td>ALcoholic drinks</td>
<td>.054</td>
</tr>
<tr>
<td>Dish detergent (tablets)</td>
<td>.066</td>
</tr>
<tr>
<td>Powdered laundry detergent</td>
<td>.096</td>
</tr>
<tr>
<td>Margarine/oil</td>
<td>.097</td>
</tr>
<tr>
<td>Liquid dish detergent</td>
<td>.136</td>
</tr>
<tr>
<td>Milk substitutes</td>
<td>.426</td>
</tr>
<tr>
<td>Sweet-and-sour canned products</td>
<td>.541</td>
</tr>
<tr>
<td>Ready-to-eat meals</td>
<td>.682</td>
</tr>
<tr>
<td>Toilet and kitchen tissues</td>
<td>.761</td>
</tr>
<tr>
<td>Pastries</td>
<td>.779</td>
</tr>
<tr>
<td>Crackers</td>
<td>.785</td>
</tr>
<tr>
<td>Canned soup</td>
<td>.877</td>
</tr>
<tr>
<td>Salty snacks</td>
<td>.945</td>
</tr>
<tr>
<td>Eggs</td>
<td>1.256</td>
</tr>
<tr>
<td>Hair care</td>
<td>1.519</td>
</tr>
<tr>
<td>Baking products</td>
<td>1.648</td>
</tr>
<tr>
<td>Cleaning materials</td>
<td>1.792</td>
</tr>
<tr>
<td>Sugar</td>
<td>1.938</td>
</tr>
<tr>
<td>Cleaners</td>
<td>2.259</td>
</tr>
<tr>
<td>Deodorant</td>
<td>3.546</td>
</tr>
</tbody>
</table>

The table reports the (posterior) learning parameter, i.e., the ratio of consumption signal variance to initial variance. Higher levels of the learning parameter imply less learning.

5.2. Stage 2: Linking consumption-based learning to household and category drivers

5.2.1. Regression coefficients

Table 5 reports the results of the second-stage regression. With an adjusted R² of 32.3%, the model fit is very good, and, taken together, the drivers explain a significant portion of the learning parameter variation (p < .001). The variance inflation factors all remain below 6, indicating that we do not have collinearity problems. For interpretation of the results, it is important to recall that our dependent variable is the negative of the posterior log-learning parameter such that variables associated with positive coefficients imply more learning. To facilitate the comparison of effects across variables, we focus on the standardized coefficients.

Zooming in on the drivers underlying incentives to learn first, we find that learning is significantly higher in categories with high performance risk (\(\beta = .094, p < .01\)) and high involvement (\(\beta = .477, p < .01\)). This finding matches expectations: consumers closely monitor their...
consumption experiences and update their quality beliefs in categories where the risk of making a wrong choice is higher and/or has personal relevance to them. We also find consumption-based learning to be stronger in more expensive categories (β = .082, p < .01). Although this effect was anticipated, it was surprisingly weak (partial eta-squared of .009, compared to .143 for involvement17), suggesting that consumers’ motivation to learn stems from non-monetary rather than monetary stakes.

As expected, variety seeking reduces the degree of consumption-based quality updating. Learning is significantly lower in categories with a high need for variety (β = −.396, p < .01) and among consumers who buy many different brands within a given category (β = −.079, p < .01). In a similar vein, consumers’ general tendency to be brand loyal exerts a positive impact on learning (β = .067, p < .01) – completing the consistent pattern of results for this dimension.

With regard to domain familiarity, we obtain a significant positive main effect for category purchase frequency (β = .391, p < .01). At the same time, the coefficient of the squared term is also significant with the opposite sign (β = −.158, p < .01). This finding suggests that households with more extensive category expertise (are) better (able to) keep track of the performance of specific brands based on the most recent consumption signals. However, this effect occurs only up to a certain point, after which it levels off, perhaps because there is little remaining uncertainty to be resolved.

Turning to the control variables, we find significantly more learning in food categories (β = .265, p < .01). Consumers who generally plan their purchases in advance are no different from others in their consumption-based learning (β = .003, p = .794). Interestingly, households that act as shopping mavens (based on the self-report survey data) appear to learn slightly less (β = −.024, p = .024) – a possible sign of overconfidence on their part (Hutchinson & Eisenstein, 2008).

Taking the findings together, we find that all three dimensions significantly contribute to the degree of learning. Whereas most of the explained variation comes from consumers’ incentives to learn (48%), variety seeking (19%) and domain familiarity (14%) also exert a non-negligible influence.18 Zooming in on the role of category- versus household-specific factors, however, it appears that the former play a far more important role, which we turn to next.

### 5.2.2. Do some households learn more than others?

A somewhat surprising outcome of the second-stage regression is that household characteristics that are independent of the category explain only a small portion of the variation in learning. This finding raises the following question: are there households that learn more (less) than others across categories, and can we profile them? To explore this issue, we proceed as follows.19 For each category, we first determine the median level of the learning parameter across households. Next, for each household, we assess the fraction of categories where it is present, in which it exhibits above-median learning (i.e., has a learning parameter below the category median). Households with a fraction of zero would be ‘weak learners’; that is, they learn less than others in all the categories that they buy. Conversely, households with a fraction of one would learn more strongly than others for all purchased items.

Fig. 1 displays the histogram of this metric. A first observation is that many households are situated in the middle, indicating that although they are relatively strong learners in half of the categories, they are revealed to be weak learners in the other half. This finding suggests that consumption-based quality learning is not simply a household trait; it fits in with the notion that consumers (i) often ‘trade down’ (go for cheap alternatives) in some categories to be able to ‘trade up’ (focus on quality) in others (Costa, Dalens, Jacobsen, & Ramsö, 2006; Nielsen, 2008; Silverstein & Butman, 2007; Steenkamp, van Heerde, & Geyskens, 2010) and (ii) update their brand quality beliefs predominantly in the latter categories, where quality is of focal importance.20

Meanwhile, Fig. 1 does point to substantial variation across households, with a non-negligible subgroup being located at the extremes of the scale. To better understand the profile of these households, we split the sample into a lower-quartile segment (fraction below .4: the weak learners), a higher-quartile segment (fraction above .6: the strong learners), and a mid-segment. We then link segment membership to the households’ socio-demographic characteristics through a multino- 

### Table 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable notation</th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>3.895</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Purchase frequency</td>
<td>CatKnow</td>
<td>0.175</td>
<td>.391</td>
<td>.000</td>
</tr>
<tr>
<td>Purchase frequency_squared</td>
<td></td>
<td>−.003</td>
<td>−.158</td>
<td>.000</td>
</tr>
<tr>
<td>Category expensiveness</td>
<td>Expensiveness</td>
<td>0.001</td>
<td>0.082</td>
<td>.000</td>
</tr>
<tr>
<td>Performance risk</td>
<td>Risk</td>
<td>0.922</td>
<td>0.094</td>
<td>.000</td>
</tr>
<tr>
<td>Involvement</td>
<td>Involvement</td>
<td>3.490</td>
<td>0.477</td>
<td></td>
</tr>
<tr>
<td>Number of brands consumed</td>
<td>Number_Brand</td>
<td>−.058</td>
<td>−.079</td>
<td>.000</td>
</tr>
<tr>
<td>Brand loyalty</td>
<td>Brand_Loyalty</td>
<td>0.219</td>
<td>0.067</td>
<td>.000</td>
</tr>
<tr>
<td>Need for variety</td>
<td>Need_Variety</td>
<td>−.273</td>
<td>−.396</td>
<td>.000</td>
</tr>
<tr>
<td>Dummy for food categories</td>
<td>Food</td>
<td>1.315</td>
<td>.265</td>
<td>.000</td>
</tr>
<tr>
<td>Mavenism</td>
<td>Mavenism</td>
<td>−.099</td>
<td>−.024</td>
<td>.002</td>
</tr>
<tr>
<td>Purchase planning</td>
<td>Planning</td>
<td>−.006</td>
<td>−.003</td>
<td>.79</td>
</tr>
</tbody>
</table>

N = 6668

R² = .570, R²_adj = .323

F test: p < .001

Note: Dependent variable: − log of learning parameter (= − log (λ²/σλ²)), weighted regression.

---

17 We note that involvement and expensiveness are not correlated in our data (correlation coefficient: −.003, p > .10), indicating that they capture separate constructs. Similarly, the correlation between performance risk and expensiveness, though significant, remains very low (correlation coefficient: .05; p < .05).

18 Figures based on the ANOVA results associated with our second-stage regression. Control factors account for the remaining portion of explained variation (13%).

19 Because most households purchase only a subset of categories, cluster analysis (i.e., grouping households based on the category-learning parameters) is not appropriate here, as it would suffer from a major missing values problem.

20 See also Erdem, Mayhew, and Sun (2001), who find use-experience-sensitive consumers to be less price-sensitive.
low-income households tend to be stronger learners: they have a significantly lower probability of being in the mid- or low-learning segment \((p < .07\) and \(p < .04,\) resp.). In contrast, the low-learning segment comprises more two-person households (rather than singles or households with children) — be it that this effect is only marginally significant \((p < .10).\) Interestingly, the age of the household head has an inverted-U effect \((p < .05),\) suggesting that the propensity to be a weak learner is lower for both young and elderly shoppers. What categories do these households learn differently about? Comparing the category-specific learning parameters, we find that weak learners fail to update their quality beliefs especially in pastries, deodorant, mouth hygiene and breakfast cereals – categories where strong learners learn significantly more. Items such as margarine, salads, and dairy drinks, where learning is strong to begin with, trigger comparable learning among the different household groups.

6. Discussion, limitations and future research

Emerging empirical evidence has already suggested that consumption-based brand-quality learning not only occurs in ‘consequential’ or evolving categories but also prevails for CPG products. However, the studies to date are restricted to only a few categories – which begs the question of generalizability – and do not examine which consumers are more inclined to learn. In this study, we explored patterns in the development of consumption-based brand-quality knowledge, across 584 households and 33 packaged goods categories, and investigated household and category characteristics that drive learning.

6.1. Findings

Our main findings are as follows. First, we obtain evidence of learning for almost all products: except in one category (where learning does not improve fit), consumers are found to update their beliefs about brand quality based on their consumption experiences. This finding underscores the importance of such learning even in mature, packaged goods settings and justifies the use of marketing tactics designed to benefit from consumption-based quality assessments over time.

Second, the degree of learning varies substantially across categories and across households within categories. In some cases, learning is minimal. This is, for instance, the case for cleaners, where even for strong-learning consumers (upper quartile), only 30% of the brand’s quality belief is shaped by the initial consumption experience. In other instances, households’ quality beliefs are predominantly shaped by recent consumption signals, as is the case for, e.g., diapers. In addition, cross-household differences are significant and substantial – in categories where learning is strong or weak overall (e.g., for toilet tissue, the inter-quartile range of the posterior learning parameter increases from a low of .79 to a high of 3.691; for deodorants: from .263 to 18.05).

Third, our results uncover category-related characteristics and household-category features that drive the degree of consumption-based learning. As expected, the need for variation and the tendency to buy several brands detract from the consumption-based updating of specific brand quality beliefs. Conversely, consumers attach far more weight to experienced consumption quality for higher-involvement CPG items and in categories with high performance risk. Higher ‘stakes’ in these categories imply that households have extra incentives to keep track of the products’ quality. Our results thus empirically corroborate that, even in CPG categories, consumer motivation is a key determinant of experience-based learning (Hoch & Deighton, 1989). Although we also observe more quality-belief updating for products where a typical purchase is expensive, this effect is much less pronounced. Therefore, non-monetary rather than monetary stakes play a pivotal role in consumers’ motivation to learn. We also find that learning is stronger in categories for which the consumer has a history of frequent purchases and thus is more familiar overall. This finding is in line with the ‘enrichment’ hypothesis, that prior knowledge reduces the cost of processing new information and enhances the ability to learn (Hoch & Deighton, 1989; Urbany et al., 1989). Following this hypothesis, frequent category buyers would be better able to interpret consumption signals and extract (brand) quality information from these signals. Although we see evidence of such an effect, we also find that it levels off for very high levels of purchase frequency – in line with the premise that encoding deteriorates at very high levels of familiarity (Johnson & Russo, 1984).

Fourth, we find some differences in consumers’ overall tendency to learn, with consumption-based updating of brand-quality beliefs being less prevalent among high-income, middle-aged, and two-person households. The overarching finding, however, is that ‘stable’ (category-independent) household characteristics explain only a small portion of the degree of learning. The vast majority of households exhibit strong learning in some categories but weak learning in others. This finding suggests not only that consumption-based quality learning is not simply a household trait but also that households tend to update their quality knowledge more or less strongly depending on the product.

### Table 6

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment 1 (weak learners)</th>
<th>Segment 2 (medium learners)</th>
<th>Segment 3 (strong learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.81**</td>
<td>2.54</td>
<td>-</td>
</tr>
<tr>
<td>Low income</td>
<td>-.663**</td>
<td>-.659**</td>
<td>-</td>
</tr>
<tr>
<td>Two-person household</td>
<td>.314**</td>
<td>.229</td>
<td>-</td>
</tr>
<tr>
<td>Age of head-of-householdb</td>
<td>-.2317**</td>
<td>-.342</td>
<td>-</td>
</tr>
<tr>
<td>Age of head-of-household_squared</td>
<td>.149**</td>
<td>.029</td>
<td>-</td>
</tr>
</tbody>
</table>

* Reference segment.

** Based on 15 subsequent age categories.

\( p < .05.\)

\( p < .10.\)
Such an implication appears in line with the premise that consumers have limited attention and cognitive abilities and closely keep track of brand quality only in product categories that have personal relevance for them (Celsi & Olson, 1988). It also appears consistent with the notion that many consumers ‘trade up and down’ (Costa et al., 2006; Nielsen, 2008; Silverstein & Butman, 2006), exhibiting a strong interest in quality (learning) in some categories while focusing more on price in others.

6.2. Implications

6.2.1. Consumer welfare

The commonly accepted view is that learning enhances consumer welfare. As indicated by Hutchinson and Eisenstein (2008), learning from experience enhances consumer expertise (they become better and more efficient in their role as consumers). Moreover, experience-based learning reduces consumers’ perception bias for specific brands, i.e., brings their quality perceptions of the brands closer to their true quality (Erdem, 1998). Therefore, learning allows consumers to select products that better match their preferences at lower cost. In addition, there exists an indirect effect through ‘supply-side’ factors. Categories in which consumers learn leave less room for ‘cheating’ and undue (misleading) quality signaling. These categories also entail stronger incentives for manufacturers to continuously monitor the quality (consistency) of their products and provide free samples, allowing consumers to assess the brand’s quality ‘firsthand’ rather than having to rely on ad messages. Therefore, learning leads to higher consumer welfare. From this perspective, it is particularly encouraging that consumers learn more in categories where they are more vulnerable or have higher stakes, that is, expensive categories, frequently bought categories, and categories featuring high performance risk. Similarly, the finding that more ‘vulnerable’ segments (low-income, large-family or single, young and elderly households) exhibit stronger learning is positive from a consumer welfare perspective.

6.2.2. Managerial implications

Our results reveal that even for CPG products, consumers continue to learn from experience with (and keep on forgetting about) incumbent brands. This finding has important implications for the managers of these brands. First, it provides a rationale for continued investments in quality, even for established brands. Although such investments may not yield an immediate boost in brand share, their effect is bound to materialize after several periods as consumers gradually update their quality perceptions across subsequent consumption occasions. A second implication is that marketing actions that stimulate trial and consumption – such as free samples – may remain powerful instruments for incumbent brands in ‘mature’ CPG categories. By highlighting categories where learning is stronger, our results guide multiproduct manufacturers in the allocation of free sampling promotions: sampling appears most promising in frequently purchased, expensive, high-involvement categories with high performance risk but may pay off less in categories where variety seeking prevails. Of course, some caution is warranted here, as the learning parameter – which captures the amount of information that consumers extract from an initial consumption experience with the brand - is not the only driver of a brand’s free trial effectiveness. First, though learning reduces quality uncertainty, it also affects consumers’ mean quality perception of the brand, and this adjustment (and the ensuing benefits from learning) is obviously more positive for brands whose true quality is high. Second, even for a given learning parameter, quality updating diminishes as consumers familiarize themselves with the brand. It follows that within the boundaries of a given category, the effectiveness of free sampling will be higher for brands that ‘deliver on their promise’ (generate truly satisfactory quality experiences) and lower for brands that already occupy a dominant market position — especially among their regular buyers.

To gain an understanding of these effects, we run a small simulation, in which we consider a manufacturer that operates in multiple categories, and focus on the different brands/categories in our analysis produced by that manufacturer. For each of these brands, we use our estimated model to dynamically simulate the effect of a free trial promotion, using our actual data set as a backdrop. The promotion ‘runs’ for one week per year, during which each consumer engaging in a category purchase receives a small version (one tenth of the regular package size) of the focal brand for free. Similarly to Erdem (1998), we assume that consumers who receive the sample actually use it. We then compare the brand shares predicted by the model with and without the sample promotion. Because these share(s) (differences) depend on the estimated parameters, which have some inherent uncertainty,

<table>
<thead>
<tr>
<th>Product category</th>
<th>Brand</th>
<th>Free trial effect</th>
<th>Category learning parameter: Mean [IQR]</th>
<th>Brand base share</th>
<th>True brand quality</th>
<th>Free trial effect compared to company average (t-statistic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canned soup</td>
<td>Brand 3</td>
<td>0.002**</td>
<td>0.000</td>
<td>0.877 [0.697–1.092]</td>
<td>0.640 +</td>
<td>–</td>
</tr>
<tr>
<td>Hair care</td>
<td>Brand 1</td>
<td>0.004‡</td>
<td>0.001</td>
<td>1.519 [1.367–1.771]</td>
<td>0.337 +</td>
<td>–</td>
</tr>
<tr>
<td>Powdered laundry detergent</td>
<td>Brand 1</td>
<td>0.006**</td>
<td>0.002</td>
<td>0.096 [0.074–1.46]</td>
<td>0.130</td>
<td>–</td>
</tr>
<tr>
<td>Flavor additives</td>
<td>Brand 2</td>
<td>0.007**</td>
<td>0.002</td>
<td>0.000 [0.000–0.000]</td>
<td>0.005</td>
<td>–</td>
</tr>
<tr>
<td>Fabric softeners</td>
<td>Brand 2</td>
<td>0.008**</td>
<td>0.003</td>
<td>0.044 [0.023–0.81]</td>
<td>0.390</td>
<td>–</td>
</tr>
<tr>
<td>Dish detergent (tablets)</td>
<td>Brand 3</td>
<td>0.010**</td>
<td>0.002</td>
<td>0.066 [0.030–241]</td>
<td>0.333</td>
<td>–</td>
</tr>
<tr>
<td>Cleaning materials</td>
<td>Brand 3</td>
<td>0.013**</td>
<td>0.004</td>
<td>1.792 [1.297–10.172]</td>
<td>0.486</td>
<td>0</td>
</tr>
<tr>
<td>Meal decanters</td>
<td>Brand 1</td>
<td>0.016**</td>
<td>0.001</td>
<td>0.020 [0.009–0.048]</td>
<td>0.361</td>
<td>–</td>
</tr>
<tr>
<td>Liquid laundry detergent</td>
<td>Brand 1</td>
<td>0.027**</td>
<td>0.007</td>
<td>0.019 [0.013–0.030]</td>
<td>0.368</td>
<td>+</td>
</tr>
</tbody>
</table>

** Effect significantly different from zero (p < 0.05).

Based on parameter estimates of the brand’s true quality relative to the leading competitors in the category (see Appendix C); + = higher than average, — = lower than average, — = far lower than average, 0 = average.

Average increase in brand share, across all (promotion and non-promotion) weeks in the data set, resulting from the free sample promotion. Increase expressed as a fraction (e.g., 0.002 refers to .2% points). In the free sample promotion week (once every 52 weeks), category buyers receive a small (10% of regular pack size) sample of the promoted brand.

IQR = inter-quartile range (see also Table 4).

For confidentiality reasons, the data provider does not allow us to reveal the manufacturer and brand names.
we repeat the simulations for 300 parameter draws from their sampling distributions and use these draws to derive the standard errors of the free sample effects.

Table 7 displays the average increase in share for the focal brands resulting from the promotion, listed from low to high. First, the free trial generates a significant positive effect for each of the brands, suggesting that such sampling indeed may play out well. Given that the promotion is ‘implemented’ only once every 52 weeks and that its impact is calculated across all (promotion and non-promotion) weeks, the effects are quite sizable. On average, the sample leads to a one percentage point higher brand share, the effects being much higher in some cases. For instance, for the company’s liquid laundry detergent brand, a share increase of 2.7 percentage points is noted — approximately 7% of the brand’s original share. These benefits will have to be weighed against the costs of the sample promotion. Second, the pattern of effects reflects the estimated category differences in degree of learning. As shown by the last column of Table 7, categories for which the share increase is significantly below the company (one percentage point) average are those with a consistently high learning parameter (low learning) (canned soup, hair care, powdered laundry detergent), whereas above-average gains are obtained for products where the learning parameter remains low (meal decorators, liquid laundry detergent). Categories where sampling does not significantly over- or underperform are those with ‘medium’ learning (fabric softeners, dish detergent tablets) or strong heterogeneity across households (cleaning materials). Therefore, our observed category differences in learning offer guidance to manufacturers in determining which products to offer free samples of. Third, notable deviations from the pattern can still be observed. Even though it belongs to the strongest-learning category (flavor additives), the free sample effect for the company brand remains relatively small (.7% points) — which may be ascribed to the fact that its true quality is far below that of leading competitors (see the true quality estimates in Appendix C). In addition, interesting reversals occur: though quality updating is stronger (i.e., the category learning parameter is lower) for canned soup than hair care, the sampling effect for canned soup than hair care, the sampling effect for relative small (.7% points) — which may be ascribed to the fact that its true quality is far below that of leading competitors (see the true quality estimates in Appendix C). In addition, interesting reversals occur: though quality updating is stronger (i.e., the category learning parameter is lower) for canned soup than hair care, the sampling effect for the company’s brand is only half, possibly as a result of its already dominant market position (64% market share, indicating that many consumers are already familiar with the brand). Therefore, as mentioned earlier, the effect of sampling hinges on the brand’s true quality and untapped market potential — something for brand managers to reckon with.

Targeted sampling may be the ideal approach in this case. Specifically, in categories with a high learning parameter, brands that already enjoy a large customer base will especially benefit from ‘competitive’ targeted promotions, i.e., offering free samples to previous non-buyers (rather than buyers) of the brand (Zhang & Wedel, 2009), who still strongly learn from the free trial experience. In a similar vein, in categories where the learning parameter itself is heterogeneous (in the illustration above, cleaning materials), the sampling effect may be boosted by targeting the stronger learners (i.e., frequent category buyers or households that buy few different brands).

6.3. Limitations and future research

Our study has limitations that open interesting opportunities for future research. First, we focused on the (four to six) top brands in each category. To the extent that leading brands are better known to consumers than less popular brands, they may be subject to weaker learning. As such, our estimate of the degree of learning is likely to be a conservative one, something to be verified in future studies. Second, our set of household and category drivers (which we needed to link with the panel households’ estimated learning parameters) was limited to the items available from the surveys. Future studies could extend this list with other perceptual and attitudinal variables. One such variable could be the ‘ambiguity’ of experience information in the product category (Hoch & Deighton, 1989), i.e., the degree to which consumers find the brand’s quality difficult to pin down because, for instance, it is easily confounded with contextual factors. Incorporating these variables could further our understanding of when and why households learn from consumption. Third, as in previous choice models with Bayesian updating, our measure of learning was an implicit one, as brand quality beliefs are unobserved. To ensure identification, we thus had to impose restrictions (and set the initial belief uncertainties and consumption signal variances to be equal across brands within a category). As indicated by Shin et al. (2012), combining observed brand choices with stated measures of brand familiarity and liking may remove the need for such restrictions and allow for a more refined assessment of learning by brand or brand-type (national brand versus private label). Finally, given our broad range of categories, we could not estimate the brand choice models simultaneously across categories, which may have dampened the cross-category effects. In addition, we focused on brand learning within categories. Analyzing how consumers update their beliefs about umbrella brands across categories is a challenging topic for future study.

Acknowledgments

The authors gratefully acknowledge AiMark for providing the data. They also thank Marnik Dekimpe, Kusum Ailawadi and Ralf van der Lans for their excellent suggestions. Finally, they thank the IJR guest-editor, Hubert Gatignon, and the anonymous members of the review team, for their very constructive and helpful comments on earlier versions of the paper.

Appendix A. Estimation issues in the first stage of the analysis

A.1. Identification

Assuming that at time $t = 0$, consumers know the brands’ mean true quality across the population and begin with rational expectations, we let $μ_{0,j} = v_{q,j}$, i.e., set the initial mean quality beliefs for a brand equal to its true quality population mean (see Crawford & Shum, 2005; Narayanan & Manchanda, 2009 for a similar approach). To obtain identification, we fix in each of the category-specific models the population mean of one of the true quality parameters to zero (i.e., $v_{q,j} = 0$ for $j = 1$). The remaining (mean) true brand qualities are identified relative to the reference brand based on the choice probability of consumers who are well experienced with a given brand and thus have low quality uncertainty for that brand. Variations within the same brand across these experienced consumers allow us to assess the standard error of the true qualities’ mixing distribution. We acknowledge that this assessment of true brand qualities rests on two assumptions. First, the consumers’ rate of forgetting should be substantially lower than their rate of learning, such that repeated consumption indeed lets their quality beliefs converge to their true qualities. Second, the observed consumption histories should be sufficiently long for experienced consumers’ beliefs to converge to the true qualities. Previous studies suggest that it typically takes consumers several periods to forget what they learned from consumption (see, e.g., Mehta et al., 2004). Moreover, our estimation period covers 2.5 years, during which consumers may accumulate experience. Therefore, it is reasonable to expect that, at least for a subset of households, we observe convergence to the true brand qualities relative to the reference brand.

The response parameters for the price and promotion variables (i.e., mean and standard error of the mixing distributions) are identified based on the fact that these variables change over time both across and within brands, with each consumer facing several changes in these variables.

The initial brand quality uncertainty, degree of learning and rate of forgetting are assessed based on consumers’ purchase patterns over time for the different brands. Like previous authors (see e.g., Mehta...
et al., 2004; Erdem & Keane, 1996), we restrict the initial variance to be equal across households and brands and set the consumption signal variance to be the same for all brands. The learning rate, i.e., the ratio of the consumption signal variance to the initial belief uncertainty, is identified based on the degree to which the (systematic) utility (and ensuing choice probability) of a brand changes following brand consumption. The initial uncertainty (variance of the prior brand beliefs) may be separately identified from the evolution of the stochastic component in the utilities following brand consumption. As shown by the mean updating equation (Eq. (4)), this random component stems from the learning errors in the consumption signals, which weigh more heavily as the initial variance increases.

Conversely, the forgetting parameter is identified based on the degree to which a brand’s choice probability changes following a period in which the brand was not consumed. Learning may be separated from forgetting by comparing the change in choice probability after consumption for occasions where the previous brand purchase occurred a long time ago or was very recent. If a consumer strongly learns but also strongly forgets, the ratio of (i) the brand’s purchase probability after a long interpurchase time, relative to (ii) the brand’s purchase probability after a short interpurchase time, will be much lower than if he hardly learns but also hardly forgets. The data allow us to separate out these two patterns based on the variation in interpurchase times for a given brand within consumers’ purchase histories.

The risk aversion parameter and the brands’ true qualities, both of which are allowed to vary across households, also must be separated out. The risk aversion parameter is not simply multiplied with the brand’s mean quality belief (which depends on the true brand quality); it also enters the utility in a separate term involving the uncertainty in this belief, which is independent of the true quality but decreases (or increases) depending on brand purchases (forgetting). Because consumers are observed to switch brands, e.g., as a result of promotions, and because purchases deterministically reduce the brand’s belief uncertainty but do not systematically increase or decrease its mean quality belief, we may separate true quality effects from risk aversion. As indicated by Crawford and Shum (2005) and Chintagunta, Jiang, and Jin (2009), the separate identification of the risk aversion parameter and the consumption signal variance is more subtle. High risk aversion goes along with a low overall degree of switching, i.e., a high overall persistence of brand choices. The consumption signal variance, in turn, is identified by the degree to which switching changes over time as a function of the number of previous brand consumptions. This pattern of changes over time is restricted by the Bayesian variance-updating expression (Eq. (5)), which allows the separation of the risk aversion and learning parameters.

A.2. Dealing with left truncation

We do not observe consumers from the beginning of their consumption history and thus face the problem of left truncation. Following previous authors (e.g., Mehta et al., 2004), we include an initialization period. Specifically, we use the first four purchases of a household to build \( \mu_{0h}^0 \) and \( \sigma_{0h}^2 \) and use these levels to initialize the estimation of the remaining data. We set the initial quality beliefs at the beginning of the initialization period equal to the mean of the true quality across the population, such that \( \mu_{0h}^0 = \mu_0 \).

A.3. Likelihood function

For any household \( h \), the likelihood of the entire purchase history in a category \( D_{Th}^h \), conditional on \( \Omega^h, E_{Th}^h \), where \( \Omega^h \) summarizes the set of (household-specific) parameters related to prior belief uncertainty, true brand qualities, sensitivity to utility determinants (i.e., price, feature/display), risk aversion, forgetting, and consumption-based learning (i.e., consumption signal variance), and where \( E_{Th}^h \) is the vector of consumption signals \( c_{jct}^h \) received by household \( h \) up to purchase occasion \( t \), is as follows:

\[
L^h(D_{Th}^h|\Omega^h, E_{Th}^h) = \prod_{t=1}^{n_h} \prod_{j=1}^{J} \text{Pr}(d^h_j = 1|\Omega^h, E_t^h)^{d^h_j}.
\]

We denote the p.d.f. of parameters in \( \Omega^h \) as \( u_{0h}(\Omega^h) \) and the p.d.f. of \( E_{Th}^h \) as \( u_{Th}(E_{Th}^h) \). The unconditional likelihood of household \( h \)’s purchase history \( D_{Th}^h \) is as follows:

\[
L^h(D_{Th}^h|\alpha_t) = \int_{\Omega^h} \int_{E_{Th}^h} L^h(D_{Th}^h|\Omega^h, E_{Th}^h) u_{0h}(\Omega^h) u_{Th}(E_{Th}^h) d\Omega^h dE_{Th}^h.
\]

Because this likelihood involves multidimensional integrals, numerical computation is prohibitively expensive. Therefore, in line with previous studies, we resort to simulated likelihood. Using \( F \) sets of scrambled Halton draws (Train, 2003) for the coefficients in \( \Omega^h \) and the consumption signals in \( E_{Th}^h \), we obtain an estimate of \( L^h \):

\[
L^h(D_{Th}^h|\alpha_t) = \frac{1}{F} \sum_{f=1}^{F} L^h(D_{Th}^h|\Omega^f, E^f).
\]

We set \( F = 100 \) because larger values are not feasible given the computational demands of the model. As a robustness check, we run the model with different sets of draws, and the results do not change.

The log-likelihood for \( N \) households is as follows:

\[
LL\left( \left\{ D_{Th}^h \right\}_{h=1}^{N} | \alpha_t \right) = \sum_{h=1}^{N} \ln L^h(D_{Th}^h|\alpha_t).
\]

We estimate the parameters in \( \alpha_t \) by maximizing this likelihood:

\[
\alpha_{\text{MLE}} = \arg \max_{\alpha_t} LL\left( \left\{ D_{Th}^h \right\}_{h=1}^{N} | \alpha_t \right).
\]

From the updating expressions, it is clear that the initial perception variance and the quality signal variance are not separately identified for each brand and consumer (only their ratio is identifiable; see Shin et al., 2012). Therefore, we must impose restrictions. One possibility is to set one element of the ratio (either the consumption signal variance, or the initial variance of the quality belief) equal to one and estimate the other. Another option is to impose homogeneity restrictions on each component, e.g., make the initial variance the same across consumers, and the consumption signal variance homogeneous across brands, an approach followed by Narayan and Manchanda (2009). We use the latter approach here. Moreover, similar to Mehta et al. (2004) and Erdem and Keane (1996), we restrict the initial variance to be homogeneous across brands to further facilitate identification.

Because the consumption signals are random draws (from a distribution with a mean equal to the true brand quality), the value of them may lift or reduce the previous mean quality belief.

---

23 From the updating expressions, it is clear that the initial perception variance and the quality signal variance are not separately identified for each brand and consumer (only their ratio is identifiable; see Shin et al., 2012). Therefore, we must impose restrictions. One possibility is to set one element of the ratio (either the consumption signal variance, or the initial variance of the quality belief) equal to one and estimate the other. Another option is to impose homogeneity restrictions on each component, e.g., make the initial variance the same across consumers, and the consumption signal variance homogeneous across brands, an approach followed by Narayan and Manchanda (2009). We use the latter approach here. Moreover, similar to Mehta et al. (2004) and Erdem and Keane (1996), we restrict the initial variance to be homogeneous across brands to further facilitate identification.

24 Because the consumption signals are random draws (from a distribution with a mean equal to the true brand quality), the value of them may lift or reduce the previous mean quality belief.

233

Appendix B. Fit of ﬁrst stage model, compared with a static model

Category

Number of brands

Flavor additives
Periodic care\diapers
Dairy products
Air refreshers
Dairy drinks
Liquid laundry detergent
Salads
Meal decorators
Breakfast cereals
Mouth hygiene
Pet food
Fabric softeners
Spices and herbs
Alcoholic drinks
Dish detergent (tablets)
Powdered laundry detergent
Margarine\oil
Liquid dish detergent
Milk substitutes
Sweet-and-sour canned products
Ready-to-eat meals
Toilet and kitchen tissues
Pastries
Crackers
Canned soup
Salty snacks
Eggs
Hair care
Baking products
Cleaning materials
Sugar
Cleaners
Deodorant

6
4
4
4
5
5
4
4
5
4
6
4
4
4
4
6
5
4
5
5
4
5
5
6
4
6
5
4
6
4
4
4
5

Static model

Bayesian learning model

Number of parameters

Loglikelihood

AIC

Number of parameters

Loglikelihood

AIC

17
13
13
13
15
15
13
13
15
13
17
13
13
13
13
17
15
13
15
15
13
15
15
17
13
17
15
13
17
13
13
13
15

1577.06
2246.74
7475.44
874.60
2914.15
1218.76
3326.95
7323.15
1638.43
1060.31
5224.73
1055.36
2162.85
3477.71
309.17
2041.71
4828.34
518.96
1267.60
1010.01
3236.55
5016.56
620.49
1958.34
1057.03
7670.29
855.84
1776.59
4538.97
808.63
1303.40
1623.11
403.39

3188.12
4519.49
14,976.89
1775.19
5858.29
2467.53
6679.89
14,672.29
3306.86
2146.62
10,483.47
2136.71
4351.70
6981.43
644.34
4117.43
9686.68
1063.91
2565.20
2050.03
6499.10
10,063.12
1270.98
3950.68
2140.06
15,374.58
1741.69
3579.18
9111.94
1643.27
2632.80
3272.22
836.79

22
18
18
18
20
20
18
18
20
18
22
18
18
18
18
22
20
18
20
20
18
20
20
22
18
22
20
18
22
18
18
18
20

1508.86
1620.99
7426.81
744.23
2860.84
1087.58
3252.14
7342.91
1506.56
914.76
4816.30
988.36
2082.12
3367.44
218.43
1199.77
4605.69
395.04
1124.12
959.40
3130.49
4643.18
556.56
1871.84
1029.64
7575.76
621.10
1628.40
4445.11
716.94
1207.41
1543.05
191.58

3061.73
3277.99
14,889.61
1524.46
5761.69
2215.16
6540.29
14,721.82
3053.13
1865.52
9676.60
2012.71
4200.24
6770.88
472.86
2443.54
9251.37
826.07
2288.24
1958.80
6296.99
9326.37
1153.11
3787.68
2095.28
15,195.52
1282.21
3292.80
8934.23
1469.89
2450.83
3122.11
423.15

Appendix C. Stage 1 estimation results for the different categories

Mean

Variance (log)

Mean

Category

Flavor additives

Periodic care
\diapers

Dairy products

Brand1
Brand2
Brand3
Brand4
Brand5
Brand6
Belief variance at t = 1
Signal st. error
Forgetting
FD
Price
Risk aversion
Brand1
Brand2
Brand3
Brand4
Brand5
Brand6
Signal st. error
Forgetting
FD
Price
Risk aversion

1.00
0.06
0.67
15.68
0.47
2.61
1.05
−16.29
−6.90
−14.41
−0.05
0.04
−4.15
−4.36
−16.29
4.88
−2.07
1.66
−0.74
3.81
0.68
−1.94
−2.25

1.00
2.25
−0.71
0.07

Fixed
⁎⁎⁎
⁎⁎⁎

1.00
−3.76
−4.28
−3.98

Fixed
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

1
−4.35
−1.15
5.12
−0.05
−0.03
0.31
3.59
0.82
−1.97

⁎⁎⁎
⁎⁎⁎
⁎
⁎⁎⁎

⁎⁎⁎
⁎⁎⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

2.18
−2.52
−0.14
0.52
0.40
−1.48
2.18
1.37
0.89
0.71

1.09
1.59
1.93
−4.24
2.17

⁎⁎⁎
⁎⁎⁎
⁎⁎
⁎⁎⁎
⁎⁎⁎

−1.03
−1.84
3.38
1.30
−5.07

Category

Meal decorators

Brand1
Brand2
Brand3

1.00
0.73
−1.03

Fixed
⁎⁎⁎
⁎⁎⁎
⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎
⁎⁎
⁎⁎

†
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

Fixed
⁎⁎⁎
⁎⁎⁎

Breakfast
cereals
1.00
0.26
1.24

Fixed

1.00
0.58
0.17

1.00
1.32
0.34
0.91

†

1.89
−3.10
−0.02
6.28
0.11
−0.19
−0.98
−1.18
−0.80
−0.13

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

−0.05
−0.07
4.15
−2.50
−3.07

†
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

Mouth hygiene

⁎⁎⁎

Air refreshers

Fixed
⁎

Fixed
⁎⁎⁎
⁎⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

⁎⁎⁎
⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

Pet food
1.00
−0.70
0.06

Dairy drinks
1.00
−4.83
−5.23
−2.83
−3.72

Fixed
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

2.32
−1.65
−2.97
0.11
4.33
−1.22
2.58
−0.97
1.12
2.08
0.85

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

0.11
1.08
4.51
5.96
−3.26

1.00
−0.51
−0.17

1.00
0.74
0.23
0.36
0.81

Fixed
⁎⁎

⁎⁎⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

1.98
−1.82
−2.72
6.20
−0.39
−0.35
−17.32
−0.74
−6.43
−0.02
−0.04

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎
⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

0.71
2.77
−0.33
−0.20
−2.50

⁎⁎
⁎⁎⁎

Fabric softeners
Fixed
⁎⁎⁎

Liquid laundry
detergent

Fixed
⁎⁎

⁎

⁎⁎⁎

Spices and herbs
1.00
3.55
−1.82

Fixed
⁎⁎⁎
⁎⁎⁎

Salads
1.00
−2.02
−2.14
−1.06

Fixed
⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

0.96
−7.13
−0.65
3.02
−0.10
−0.61
0.83
−0.92
−2.43
0.16

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

1.02
1.05
−0.97
−1.33
−4.89

†
⁎⁎⁎
⁎⁎⁎
†
⁎⁎

⁎⁎⁎
⁎⁎⁎
⁎⁎⁎

Alcoholic
drinks
1.00
−8.46
−9.21

Fixed
⁎⁎⁎
⁎⁎⁎

(continued on next page)


## Appendix C (continued)

<table>
<thead>
<tr>
<th>Category</th>
<th>Meal decorations</th>
<th>Breakfast cereals</th>
<th>Mouth hygiene</th>
<th>Pet food</th>
<th>Fabric softeners</th>
<th>Spices and herbs</th>
<th>Alcoholic drinks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand4</td>
<td>0.83 **</td>
<td>0.82 ***</td>
<td>−0.13</td>
<td>−0.53</td>
<td>−0.48 *</td>
<td>−0.40</td>
<td>−8.18 ***</td>
</tr>
<tr>
<td>Brand5</td>
<td>−0.17</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand6</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Belief variance at ( t = 1 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal st. error</td>
<td>1.06 **</td>
<td>2.61 ***</td>
<td>2.60 ***</td>
<td>2.70 ***</td>
<td>1.00 ***</td>
<td>2.45 ***</td>
<td>4.54 ***</td>
</tr>
<tr>
<td>Forgetting</td>
<td>−1.87</td>
<td>−1.09 ***</td>
<td>−1.08</td>
<td>−0.51</td>
<td>−4.00</td>
<td>−0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>FD</td>
<td>2.43 ***</td>
<td>5.23 ***</td>
<td>4.58 ***</td>
<td>3.18 ***</td>
<td>5.09 ***</td>
<td>−1.47 ***</td>
<td>0.52 ***</td>
</tr>
<tr>
<td>Price</td>
<td>−0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.27 ***</td>
<td>−0.16</td>
<td>−0.02</td>
<td>−0.88 ***</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−0.62 ***</td>
<td>−0.53 ***</td>
<td>−0.57 **</td>
<td>−0.80 ***</td>
<td>−0.29</td>
<td>−0.87 ***</td>
<td>−2.13 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance [log]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand1</td>
</tr>
<tr>
<td>Brand2</td>
</tr>
<tr>
<td>Brand3</td>
</tr>
<tr>
<td>Brand4</td>
</tr>
<tr>
<td>Brand5</td>
</tr>
<tr>
<td>Brand6</td>
</tr>
<tr>
<td>Signal st. error</td>
</tr>
<tr>
<td>Forgetting</td>
</tr>
<tr>
<td>FD</td>
</tr>
<tr>
<td>Price</td>
</tr>
<tr>
<td>Risk aversion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Dish detergent (tablets)</th>
<th>Powdered laundry detergent</th>
<th>Margarine/oil</th>
<th>Liquid dish detergent</th>
<th>Milk substitutes</th>
<th>Sweet-and-sour</th>
<th>canned products</th>
<th>Ready-to-eat meals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand1</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
</tr>
<tr>
<td>Brand2</td>
<td>1.41 *</td>
<td>0.86</td>
<td>−1.25 *</td>
<td>0.27 *</td>
<td>0.39 *</td>
<td>1.38 *</td>
<td>−0.98 *</td>
<td></td>
</tr>
<tr>
<td>Brand3</td>
<td>0.45</td>
<td>0.90</td>
<td>−0.73</td>
<td>−0.12</td>
<td>0.03</td>
<td>−0.85</td>
<td>−1.19</td>
<td></td>
</tr>
<tr>
<td>Brand4</td>
<td>1.25 **</td>
<td>1.64 ***</td>
<td>−0.21 *</td>
<td>1.62 ***</td>
<td>−0.42 *</td>
<td>1.53 *</td>
<td>−1.09 **</td>
<td></td>
</tr>
<tr>
<td>Brand5</td>
<td>1.02</td>
<td>0.44 ***</td>
<td></td>
<td></td>
<td></td>
<td>0.40</td>
<td>−1.27 ***</td>
<td></td>
</tr>
<tr>
<td>Brand6</td>
<td>1.09 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Belief variance at ( t = 1 )</strong></td>
<td>2.71 **</td>
<td>1.42</td>
<td>1.60 ***</td>
<td>1.00 ***</td>
<td>1.07 ***</td>
<td>2.14 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal st. error</td>
<td>−1.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forgetting</td>
<td>−2.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td>10.22 ***</td>
<td>5.65 ***</td>
<td>0.37</td>
<td>5.10 ***</td>
<td>3.78 ***</td>
<td>4.48 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>0.07</td>
<td>0.00</td>
<td>−0.75 ***</td>
<td>−0.40 ***</td>
<td>−0.83 ***</td>
<td>0.78</td>
<td>−0.06</td>
<td></td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance [log]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand1</td>
</tr>
<tr>
<td>Brand2</td>
</tr>
<tr>
<td>Brand3</td>
</tr>
<tr>
<td>Brand4</td>
</tr>
<tr>
<td>Brand5</td>
</tr>
<tr>
<td>Brand6</td>
</tr>
<tr>
<td>Signal st. error</td>
</tr>
<tr>
<td>Forgetting</td>
</tr>
<tr>
<td>FD</td>
</tr>
<tr>
<td>Price</td>
</tr>
<tr>
<td>Risk aversion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Baking products</th>
<th>Cleaning materials</th>
<th>Sugar</th>
<th>Cleaners</th>
<th>Deodorant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand1</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
<td>1.00</td>
<td>1.00 Fixed</td>
<td>1.00 Fixed</td>
</tr>
<tr>
<td>Brand2</td>
<td>1.59 ***</td>
<td>3.37 ***</td>
<td>−0.14</td>
<td>0.43 *</td>
<td>1.31</td>
</tr>
<tr>
<td>Brand3</td>
<td>−0.83 ***</td>
<td>1.78 ***</td>
<td>0.43</td>
<td>1.31 ***</td>
<td>7.13</td>
</tr>
<tr>
<td>Brand4</td>
<td>−0.66 ***</td>
<td>0.38 ***</td>
<td>−0.11</td>
<td>1.43 ***</td>
<td>7.45</td>
</tr>
<tr>
<td>Brand5</td>
<td>−1.38 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand6</td>
<td>−1.40 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Belief variance at ( t = 1 )</strong></td>
<td>3.41 ***</td>
<td>2.54 ***</td>
<td>2.35 ***</td>
<td>1.20 ***</td>
<td>1.05</td>
</tr>
<tr>
<td>Signal st. error</td>
<td>1.60</td>
<td>−1.11 ***</td>
<td>1.26 ***</td>
<td>0.97 ***</td>
<td>−9.97</td>
</tr>
<tr>
<td>Forgetting</td>
<td>−33.77</td>
<td>−2.78 ***</td>
<td>−14.94</td>
<td>−17.04</td>
<td>−1.50</td>
</tr>
<tr>
<td>FD</td>
<td>2.43 ***</td>
<td>3.29 ***</td>
<td>−0.19</td>
<td>5.58 ***</td>
<td>0.50</td>
</tr>
<tr>
<td>Price</td>
<td>−0.11</td>
<td>−0.97 ***</td>
<td>−0.59 ***</td>
<td>−0.05 ***</td>
<td>0.81 ***</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−1.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance [log]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand1</td>
</tr>
<tr>
<td>Brand2</td>
</tr>
<tr>
<td>Brand3</td>
</tr>
<tr>
<td>Brand4</td>
</tr>
<tr>
<td>Brand5</td>
</tr>
<tr>
<td>Brand6</td>
</tr>
<tr>
<td>Signal st. error</td>
</tr>
<tr>
<td>Forgetting</td>
</tr>
<tr>
<td>FD</td>
</tr>
<tr>
<td>Price</td>
</tr>
<tr>
<td>Risk aversion</td>
</tr>
</tbody>
</table>

\(^* \) p < .1, **p < .05, ***p < .01, ****p < .001. FD = feature display.
Social interactions in customer churn decisions: The impact of relationship directionality

Michael Haenlein *
Marketing Group at ESCP Europe, 79 Avenue de la République, F-75011 Paris, France

**Article Info**

Article history:
First received in 10, February 2013 and was under review for 10 months. Available online 19 April 2013
Area Editor: Werner J. Reinartz

Keywords:
Customer retention
Social networks
Directed network
Undirected network
Stratified Cox Proportional Hazards model

**Abstract**

The impact of social factors on individual-level decision making has been a subject of general interest within the marketing field. However, studies analyzing social interactions and social contagion have, to a great extent, focused on the importance of social interactions in the customer acquisition process and have relied on the use of undirected networks. Our study contributes to the literature stream by focusing on two elements that have been analyzed less frequently. Specifically, we focus on the importance of social interactions in the customer retention process within a directed social network. Using the customer base of a mobile phone provider, we rely on call detail records to investigate the churn behavior of 3431 focal actors. We provide evidence for social interactions in customer churn decisions and show that, at any given point in time, a focal actor is significantly and substantially more likely to defect from a provider if other individuals to whom that actor is socially connected have previously defected from the provider. However, this effect is limited to social contacts with whom the focal actor has outgoing calling relationships and who have churned relatively recently (in our sample, less than 5 weeks prior to the point in time that is under examination). We therefore provide empirical evidence demonstrating that social effects do play a role in customer retention but only when tie directionality and churn recency are taken into account.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decades, it has been widely accepted within the marketing field that consumer behavior cannot be fully understood when individual actors are studied in isolation and that social factors play an important role in shaping individual-level decision making. This perspective is particularly prevalent in the analysis of customer acquisition strategies, specifically in the area of new product adoption. Studies in those contexts have argued that socially influenced decisions occur in situations in which two or more actors make a purchase or consumption decision jointly (Hartmann, 2010) or as a result of social contagion, i.e., following exposure to other consumers’ knowledge, attitudes or behaviors concerning a product (Iyengar, Van den Bulte, & Valente, 2011). Variables that measure the aspects of social exposure have therefore gained importance in models of innovation diffusion (Libai, Muller, & Peres, 2013; Nair, Manchanda, & Bhatia, 2010).

The vast majority of this research has focused on the importance of social interactions in the customer acquisition process and has relied on the use of undirected networks, i.e., networks that consider the connections between individuals without taking into account who initiates each connection. Our study contributes to the literature by looking at two elements that have been analyzed less frequently. Specifically, we investigate the importance of social interactions in the customer retention process within a directed social network, i.e., a network in which the initiator of each connection is identified. Our analysis uses the call history database of a mobile phone operator to analyze the social interactions in customer churn decisions. Specifically, we look at situations in which two or more people who are socially connected decide to leave a company, either simultaneously or in close temporal proximity to one another. Our approach is consistent with a larger body of research that relies on electronically recorded communication patterns to approximate social relationships within a larger network (e.g., Goldenberg, Han, Lehmann, & Hong, 2009; Hill, Provost, & Volinsky, 2006; Nitzan & Libai, 2011; Trusov, Bucklin, & Pauwels, 2009). For our analysis, we used a directed network that allowed us to differentiate between the impact of churn among outgoing calling relationships (i.e., a focal actor calls another actor in the network or an alter) and among incoming calling relationships (i.e., an alter calls a focal actor).

Our results show that, at any given point in time, a focal actor is significantly and substantially more likely to defect from a provider if other individuals to whom that actor is socially connected have previously defected from the provider. However, this effect is limited to social contacts with whom the focal actor has outgoing calling relationships and who churned relatively recently (in our sample, less than 5 weeks prior to the point in time that is under examination). We also show that churn among a focal actor’s social contacts has a nonlinear effect on the hazard that the focal actor will defect as well.

The author would like to thank Barak Libai, Thomas Valente and Christophe Van den Bulte for their helpful comments and discussion. * Tel.: +33 1 49 23 26 02. E-mail address: haenlein@esceurope.eu.

0167-8116/$ – see front matter © 2013 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.ijresmar.2013.03.003
specifically, a focal actor’s hazard of defection increases quadratically with the number of recent churn events among his or her outgoing calling relationships.

2. Contribution compared to prior research

By focusing on the importance of social interactions in the customer retention process within a directed social network, our work provides the following two contributions compared to previously published work: First, prior marketing literature on the importance of social influence has focused extensively on customer acquisition, specifically on new product adoption (e.g., Rogers, 2003; Van den Bulte & Wyts, 2007; Wyts, Dekimpe, Gijsbrechts, & Pieters, 2010). Relatively few studies have looked at the impact of social networks on consumer behavior once customers have been acquired, i.e., behavior that is contingent on an initial product or brand choice decision. Some notable exceptions include the work of Haenlein (2011), who looks into the social network effects in post-acquisition product usage and customer-level revenue and Nitzan and Libai (2011), who analyze social contagion in the customer retention process.

The lack of research regarding social influence on customer retention is surprising, because customer acquisition and retention initiation are only one part of a firm’s CRM process (Reinartz, Krafft, & Hoyer, 2004), and the analysis of factors influencing retention, relationship duration, and customer churn is an active area of research (e.g., Bolton, 1998; Lemmens & Croux, 2006; Neslin, Gupta, Kamakura, Lu, & Mason, 2006; Schweidel, Fader, & Bradlow, 2008). In addition, studies that have been conducted outside of the marketing field provide clear indications that social effects play an important role not only in initiation but also in termination decisions. Medical research, for example, has highlighted the importance of social effects in the cessation of smoking (Christakis & Fowler, 2008) and of heroin and cocaine consumption (Buchanan & Latkina, 2008). Likewise, studies on human resource management have stressed the importance of communication networks in the context of collective turnover decisions (Bartunek, Huang, & Walsh, 2008; Castilla, 2005; Krackhardt & Porter, 1986). Similar effects have been found in studies of decisions regarding deserting the army (De Paula, 2009) and marriage dissolution (McDermott, Fowler, & Christakis, 2009). In light of these findings, it seems likely that social effects should also influence a customer’s decision to remain loyal to a certain company, and, indeed, Nitzan and Libai (2011) have provided empirical evidence that this is the case. Our study continues this stream of research by providing further insight into the dynamics that underlie social influence in the customer retention process.

Second, we contribute to the prior literature by analyzing social influence within a directed (versus an undirected) network. Across a large range of disciplines, including organizational science, finance, psychology and marketing, studies that analyze social networks have focused nearly exclusively on undirected networks. In marketing specifically, this is true for analytical (Zubcsek & Sarvary, 2011) and empirical research (Trusov, Bodapati, & Bucklin, 2010) and applies to social interactions in processes of customer acquisition (Nair et al., 2010) as well as retention (Nitzan & Libai, 2011). In fact, an analysis of the journals that are covered in the EBSCO database reveals only a handful of articles that rely on directed social networks and that specifically compare the difference in influence between the number of incoming connections (indegree) and the number of outgoing connections (outdegree). See Table 1 for an overview of these studies.

Interestingly, the studies that have focused on directed networks reveal that while the impact of incoming and outgoing relationships is comparable in some settings, it is not the case in all situations. This empirical finding can be explained by the fact that, depending on the situation, incoming and outgoing relationships can have fundamentally different meanings. For example, considering another person as a friend is not necessarily identical to being considered as a friend, and analyzing both types of relationships in the same manner carries the risk of confounding different effects. Empirical evidence for this finding can be found in Christakis and Fowler (2007), who analyze the spread of obesity within a social network and show that results differ depending on whether A perceives B as a friend, B perceives A as a friend, or both A and B perceive each other as friends.

An additional potential benefit of analyses that rely on directed networks is that they entail fewer identification issues and reflection biases compared with analyses of undirected networks, and therefore they can provide stronger evidence for social contagion. Within an undirected network, it is sometimes difficult to establish whether correlation in behavior is due to social contagion or to other factors, such as endogenous group formation, correlated unobservable variables or simultaneity (Manski, 1993; Moffitt, 2001). Within a directed network, it is possible to compare the impact of incoming versus outgoing relationships. If the two types of relationships differ, for example, in their influence on the time of adoption or time to churn, it can be seen as stronger evidence for social effects (Bramoulli, Djebbari, & Fortin, 2009). Christakis and Fowler (2007), for example, used this approach to rule out the possibility that the spread of obesity is driven by unobserved environmental factors rather than by social contagion. Our study builds on this line of thinking by comparing the differential effects of incoming versus outgoing relationships on the retention process.

3. Relationship directionality and social interactions in customer churn decisions

The occurrence of social influence in customer churn decisions may stem from a variety of factors. On the one hand, the decision of a customer to leave her current supplier in response to the departure of one of her social contacts might be driven by economic or utilitarian considerations, often resulting from naturally occurring or artificially induced network externality effects. In the context of the cellular phone industry, for example, which corresponds to our empirical setting, companies frequently charge different prices for in-network versus out-of-network traffic, and it has been shown that such differential pricing influences customers’ choices of supplier as well as their subsequent usage behavior (Gerpott, 2008; Hoernig, 2007; Staahl Gabrielsen & Vагstad, 2008). On the other hand, social influence in customer churn decisions might result from social factors. The churn of a friend can, for example, lead to the loss of certain social benefits that arise from patronizing the same company, which is consistent with what has been discussed from the perspective of social enrichment (Fernandez, Castilla, & Moore, 2000; Schmitt, Skiera, & Van den Bulte, 2011) and balance theory (Krackhardt & Porter, 1985; Nitzan & Libai, 2011). In most real-life situations, economic and/or utilitarian factors and social factors can be expected to influence decision making simultaneously.

The use of a directed network to analyze social interactions in customer churn decisions can provide valuable insights regarding the drivers of such decisions. It seems likely that the impact of churn among incoming and outgoing calling relationships is not identical and that relationship directionality plays an important moderating role. Two arguments can be used to justify such thinking: First, churn among incoming versus outgoing calling relationships can be associated with different sets of economic and/or utilitarian and social consequences for the focal actor. Looking at the economic and/or utilitarian reasons, a churn event among outgoing calling relationships can represent a financial burden for a focal actor, which may increase the hazard that the focal actor will defect. This is the case when the cellular service provider follows the “calling party pays” principle, according to which the mobile phone subscriber does not pay for incoming calls but only pays for outgoing calls (Valletti & Houpis, 2005) and when out-of-network calls are priced higher than in-network calls. In such a situation, the churn of a party from whom a focal actor receives only incoming calls has no financial consequences for the focal actor, as receiving calls does not generate any cost, regardless of whether those calls stem from within
the provider's network or from a competing network. The situation is different for outgoing calls, as a focal actor incurs higher costs following the churn of a person whom he or she calls, as communication is subsequently priced according to out-of-network fees rather than in-network fees, which are generally cheaper.¹

Regarding the social factors, outgoing calling relationships are purposefully initiated by a focal actor. While people cannot necessarily control who calls them, they have full control over the calls that they place themselves. In placing a call to another person, a focal actor essentially signals that the person is “worth the effort” of a call (versus waiting until the other person calls of his or her own accord) and therefore might be more important to the focal actor compared with the social contacts from whom the focal actor only receives calls. Outgoing calling relationships therefore have a different meaning attached to them in comparison to incoming ones and represent links to others who might have a more persuasive influence on the focal person (Hall & Valente, 2007). If this is indeed the case, a focal actor should be more likely to follow suit after the defection of an outgoing calling relationship as opposed to an incoming one.

Further, one should note that a directed network constitutes a more accurate representation of reality compared with an undirected network. Many of the processes that underlie the formation of social networks are directional in nature. Networks of referral (Reingen & Kernan, 1986) or advice (Lazega & Duijn, 1997), for example, build on notions of status and prestige. These concepts are related to the number of incoming relationships or indegree, i.e., the number of people who seek information from A, but not the number of outgoing relationships or outdegree, i.e., the number of people from whom A seeks information (Iyengar et al., 2011; Van den Bulte & Wuyts, 2007). Models that are designed to describe the formation of such networks, such as exponential random graph (p*) models (Holland & Leinhardt, 1981) or the Jackson–Rogers approach (Jackson & Rogers, 2007) therefore take explicit account of the direction of a relationship. Transforming an inherently directed network into an undirected one involves the addition of certain links that have were not initially present (e.g., A → B becomes A–B, which in turn is equivalent to A → B and B → A). This transformation represents a distortion of the underlying social network structure, which could bias subsequent analyses that are based on an undirected network.

### 4. Overview of the main constructs and variable definitions

Our study investigates social interactions in customer churn decisions in the context of the cellular phone industry and specifically the different effects of churn among incoming versus outgoing calling relationships. For the purpose of our analysis, we define churn as the termination of an existing service agreement between a customer and his or her cellular phone company. Our analysis focuses only on customers who maintain a contractual service agreement with their cellular phone companies and, therefore, excludes non-contractual relationships (prepaid mobile phones). We differentiate between two types of churn – voluntary churn (i.e., cases in which the decision to terminate the service agreement is made by the customer) and involuntary churn (i.e., cases in which the decision is made by the cellular phone provider) – because different causes of churn have been shown to evolve differently over time (Braun & Schweidel, 2011).

We focus our analysis on situations in which two or more customers between whom there is a calling relationship – referred to herein as social contacts – decide to leave a company either simultaneously or in close temporal proximity to one another. We define three types of calling relationships: focal customer A receives calls from customer B (an incoming calling relationship for person A); focal customer A calls customer B (an outgoing calling relationship for person A); and customers A and B call each other (a mutual calling relationship for person A). The relationship types are not mutually exclusive; e.g., two customers who have a mutual calling relationship with each other also have incoming and outgoing calling relationships with each other. We limit our analysis to first-order effects, i.e., relationships between a focal actor and the other actors who either call or are called by the focal actor directly. Potential higher-order effects (e.g., the impact of

### Table 1

Selective review of papers investigating influence within directed networks.

<table>
<thead>
<tr>
<th>Source</th>
<th>Network type</th>
<th>Indegree</th>
<th>Outdegree</th>
<th>Dependent variable</th>
<th>Key finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haenscheid and Beckman</td>
<td>4960 medium- and large-sized firms in four industry sectors</td>
<td>Number of outside directors who sit on the focal firm's board</td>
<td>Number of directors from the focal firm who sit on an outside board</td>
<td>Number of acquisitions completed by the focal firm</td>
<td>Sig. positive impact of outdegree; Impact of indegree not sig.</td>
</tr>
<tr>
<td>Hansen (1999, 2002)</td>
<td>41 divisions of a multinational electronics and computer company</td>
<td>Number of divisions that nominate focal division as a source of advice</td>
<td>Number of other divisions that the focal division nominates as a source of advice</td>
<td>Project completion time</td>
<td>Hansen (1999): Impact of indegree and outdegree not sig. Hansen (2002): Sig. negative impact of indegree and outdegree Fund performance: Sig. positive impact of indegree and outdegree; Exit rate: Sig. positive impact of indegree</td>
</tr>
<tr>
<td>Hochberg, Ljungqvist, and Lu (2007)</td>
<td>3469 venture capital (VC) funds headquartered in the United States</td>
<td>Number of times a VC firm is invited to co-invest in other VC firms' deals</td>
<td>Number of other VCs a VC firm has invited into its own syndicates</td>
<td>Fund performance and Exit rate</td>
<td>Sig. negative impact of outdegree; Impact of indegree not sig.</td>
</tr>
<tr>
<td>Mercer and Derosier (2010)</td>
<td>1016 third grade students in 11 schools</td>
<td>Number of received friendship nominations in fall (October)</td>
<td>Number of nominations given by a particular physician to other physicians for referral or discussion</td>
<td>Number of new friendship nominations received in spring (April)</td>
<td>Adoption and post-adoption usage of newly launched prescription drug</td>
</tr>
<tr>
<td>Iyengar et al. (2011)</td>
<td>193 physicians in three US cities</td>
<td>Number of nominations</td>
<td>Fund performance and Exit rate</td>
<td>Fund performance and Exit rate</td>
<td>Sig. negative impact of outdegree; Impact of indegree not sig.</td>
</tr>
</tbody>
</table>

Note: Based on all of the articles that were covered in the EBSCO database (Academic Search Premier, Alt Health Watch, Business Source, Business Source Premier, Communication & Mass Media Complete, EconLit, Health Source: Nursing/Academic Edition, MEDLINE, PsycARTICLES, Psychology and Behavioral Sciences Collection, PsycINFO, Sociological Collection) where the full text includes the expressions “social network,” “indegree” and “outdegree.” Excludes conceptual articles, descriptive analyses, manuscripts that do not report separate results for indegree and outdegree, conference proceedings and papers not written in English language.

¹ To test the difference in pricing between in-network and out-of-network calls in our sample, we performed a regression analysis in which we explained the amount of money that was paid by a customer through the number of outgoing minutes that were spent in-network and out-of-network. Our results indicate that after taking account of a series of dummy variables that indicated service plan choice, out-of-network calls are 2 to 2.5 times more expensive than in-network calls.
churn of actor C in a calling relationship in which person A calls person B and person B calls person C) are beyond the scope of our analysis.

To operationalize the extent of churn among incoming and outgoing calling relationships, we used a two-step approach: First, we collected data on all of the incoming and outgoing calls that were made by a sample of focal actors over a two-month period (time period 1). Based on this information and following the work of Onnela et al. (2007) and Nitzan and Libai (2011), we define two types of tie strength, measuring the strength of the incoming or outgoing calling relationship between focal actor i and alter j, who is also a customer of the provider, as follows:

\[
tie_{\text{incoming}}_{ij} = \frac{\text{Incoming communication volume between } i \text{ and } j}{\text{Total incoming communication volume for } i}
\]

(1)

\[
tie_{\text{outgoing}}_{ij} = \frac{\text{Outgoing communication volume between } i \text{ and } j}{\text{Total outgoing communication volume for } i}
\]

(2)

The incoming (or outgoing) communication volume between focal actor i and alter j represents the total number of minutes that were spent by focal actor i on incoming calls from customer j (or on outgoing calls to customer j) within time period 1. The total incoming and outgoing communication volume for focal actor i is the total number of minutes of all of the calls that were placed or received by i within time period 1, including calls to and from the customers of the focal provider, customers of other cellular providers, and landlines. In the case of mutual calling relationships, we determine tie strength separately for the incoming and outgoing calling relationship components. The separate determination of tie strength allows us to disentangle the impact of the strength of the incoming relationship between focal actor A and customer B (i.e., the share of person B among all of the calls that were received by actor A) from that of the strength of the outgoing relationship between the two parties (i.e., the share of person B among all of the calls that were made by actor A).

We subsequently collected data on the churn behavior of all of the focal actors as well as on the other customers with whom they maintained incoming and/or outgoing calling relationships over a second period of six months (time period 2). Based on this information, we determined two time-dependent variables for each focal actor that measure the cumulative impact of churn events that focal actor i was exposed to among his or her incoming and outgoing calling relationships, between the start of time period 2 (t0) and time t. These variables are defined in the following way:

\[
s_{in,i}(t) = \sum_{j} \sum_{t_{churn}}^{t} c_{ij} w(t - t_{churn}) tie_{\text{incoming}}_{ij}
\]

(3)

\[
s_{out,i}(t) = \sum_{j} \sum_{t_{churn}}^{t} c_{ij} w(t - t_{churn}) tie_{\text{outgoing}}_{ij}
\]

(4)

where \(c_{ij}\) is an indicator variable that equals one if customer j churned prior to or at t and zero otherwise; \(t_{churn}\) is the time at which customer j churned; and \(w()\) (0 ≤ \(w()\) ≤ 1) is a weighting function that weights churn events according to the difference between \(t\) and \(t_{churn}\).

Our definition of \(s_{in,i}(t)\) and \(s_{out,i}(t)\) is based on two assumptions: First, we expect the impact of churn among incoming and outgoing calling relationships on focal actor churn to be stronger with increasing levels of tie strength. This assumption is consistent with the observation that individuals are more affected by others with whom they have closer relationships (Brown & Reingen, 1987). It is also consistent with the research on referral programs, which has shown that referral behavior among social contacts is influenced by tie strength (Ryu & Feick, 2007). We account for this effect in the definition of \(s_{in,i}(t)\) and \(s_{out,i}(t)\) by summing over the product of the churn indicator (\(c_{ij}\)) and the tie strength instead of simply summing over \(c_{ij}\), which would be equivalent to weighting all of the churn events identically.

Second, we expect that churn events among incoming or outgoing calling relationships that happened a longer time ago should have a lesser impact on focal actor churn rates compared with churn events that occurred relatively recently. Such an assumption is consistent with the observation that the effect of word-of-mouth decays exponentially over time for both customer acquisition (Strang & Tuma, 1993; Trusov et al., 2009) and customer retention (Nitzan & Libai, 2011). This is reflected in Eqs. (3) and (4) by requiring that the weighting function \(w()\) decreases with an increase in the difference between \(t\) and \(t_{churn}\). Combined with the fact that \(w()\) should be bound between zero and one, this condition implies that \(w(0) = 1\) and \(\lim_{t \to \infty} w(t - t_{churn}) = 0\). For our analysis, we choose a weighting function that is based on the Gompertz function and that can accommodate a large variety of different shapes, ranging from the equal weighting of all churn events at one end to exponential decay at the other (see Appendix A for a more elaborate explanation):

\[
w(x) = \frac{1}{1 - e^{-\alpha x}} \left(1 - e^{\alpha x}\right) \text{ with } \alpha, \beta > 0.
\]

(5)

Using such a weighting function results in the following operationalization of \(s_{in,i}(t)\) and \(s_{out,i}(t)\):

\[
s_{in,i}(t) = \sum_{j} \sum_{t_{churn}}^{t} c_{ij} w(t - t_{churn}) tie_{\text{incoming}}_{ij}
\]

(6)

\[
s_{out,i}(t) = \sum_{j} \sum_{t_{churn}}^{t} c_{ij} w(t - t_{churn}) tie_{\text{outgoing}}_{ij}
\]

(7)

See Appendix B for a numerical example that illustrates the calculation of \(s_{in,i}(t)\) and \(s_{out,i}(t)\).

5. Research model

To test the extent to which churn among incoming and outgoing calling relationships influences focal actor churn, we estimated a stratified Cox Proportional Hazards model in which we included the two previously defined social impact measures, \(s_{in,i}(t)\) and \(s_{out,i}(t)\), as time-dependent covariates.

Hazard rate models are the method of choice when dealing with right-censored duration times (Helsken & Schmittle, 1993). Taking account of potential censoring is particularly important when analyzing social influence because truncation can trick researchers into believing that they have found evidence of social contagion when there is actually none (Van den Bulte & Jørgen, 2011). The dependent variable in a hazard model is the hazard rate \(h(t)\), which describes the conditional probability that an event (e.g., churn) will occur at time \(t\), given that it did not occur prior to \(t\):

\[
h(t) = \frac{f(t)}{1 - F(t)}
\]

(8)

where \(f(t)\) and \(F(t)\) describe, respectively, the probability density function and the cumulative density function of the duration time process.

A (semi-parametric) Cox Proportional Hazards model (Cox, 1972) assumes that \(h(t \mid x_0)\) – the hazard rate of an event at time \(t\) for actor i, who is characterized by a vector of covariates \(x_0\) – can be expressed as the product of an unspecified baseline hazard function \(h_0(t)\) at time \(t\), which is identical for all individuals, and the relative risk or hazard ratio \(exp\gamma x_0\). We use a variation on the traditional Cox Proportional Hazards model, the stratified Cox Proportional Hazards model (Lunn & McNeil, 1995; Wei, Lin, & Weissfeld, 1989), which allows for the incorporation of multiple unspecified baseline hazard functions, which are associated with different types of events. This model enables us to...
analyze the two competing risks of voluntary and involuntary churn separately, using the following functional form:

\[ h(t|x_i) = [j h_1(t) + (1 – j) h_2(t)]e^{\gamma x_i}, \tag{9} \]

where \( j \) is an indicator variable that equals one when churn is voluntary and zero otherwise, and \( h_1(t) \) and \( h_2(t) \) are the baseline hazard functions for voluntary churn and involuntary churn, respectively.

As highlighted above, working with a directed network protects us to some extent from identification issues and reflection bias, which are a common problem in the analysis of social influence (Manski, 1993; Moffitt, 2001). The inclusion of a non-parametric baseline hazard in our model also protects against identification issues (Lin & Wei, 1989). To minimize such issues even further, we include three types of control variables in our model: First, we account for the fact that people prefer to interact with others who are similar to themselves (i.e., homophily, see McPherson, Smith-Lovin, & Cook, 2001; Reagans, 2005), suggesting that people who share similar characteristics – and in particular similar consumption patterns – are more likely to be socially connected to one another. For this factor, we obtained information on actors’ observable characteristics, including basic demographics (i.e., age, gender, marital status and socio-demographics (i.e., membership in one of 13 socio-demographic groups), and included them as time-independent control variables in our model.

Second, we controlled for correlated unobservable variables that could cause similarities in behavior among social contacts. Several types of correlated unobservable variables have been discussed in the social network literature, including sales force activities (Nair et al., 2010), advertising and detailing (Van den Bulte & Lilien, 2001) and (geographical) neighborhood effects (Moffitt, 2001). To control for bias that arises from correlated unobservable variables, we obtained information on the geographic location (ZIP-code) of each actor within our sample. Given that our sample included a substantial number of different ZIP-codes, we grouped adjacent codes into larger areas. We subsequently included a full set of area-specific fixed effects in our model to take account of the similarities in exposure to marketing efforts, network quality, preferences, and other effects that result from geographical proximity.

Third, we take account of other factors that have traditionally been shown to influence individual-level churn behavior, specifically, acquisition channel and contract characteristics (i.e., service plan and contract duration). Prior research has shown the influence of acquisition channel on loyalty and cross-buying (Verhoef & Donkers, 2005), the interrelationship between current and future usage levels (Bolton & Lemon, 1999), and the duration-dependence of churn and contraction rates (e.g., Fader & Hardie, 2007; Schweidel et al., 2008). To account for the potential influence of these variables, we included an additional series of time-independent dummy variables that indicate acquisition channel and service plan choice. Contract duration was included as a time-dependent covariate and operationalized by the number of months that elapsed since the start of the contract.

The considerations that are outlined above lead to the following specification of the covariate vector \( x_i \) that is included in Eq. (9):

\[ x_i = \gamma_1 z_1 + \gamma_2 d(t) + \gamma_3 s_{in}(t) + \gamma_4 s_{out}(t) + \gamma_5 s_{in}(t)^2 + \gamma_6 s_{out}(t)^2, \tag{10} \]

where \( z_1 \) is a vector of time-independent control variables (comprising age, gender, marital status, socio-demographic group membership, geographical area, acquisition channel and service plan); \( d(t) \) is a time-dependent covariate that measures contract duration at time \( t \); and \( s_{in}(t) \) and \( s_{out}(t) \) are the social impact measures that are defined above for churn among the incoming and outgoing calling relationships.

Note that Eq. (9) allows for a quadratic influence of the two social impact measures \( s_{in}(t) \) and \( s_{out}(t) \), which enables us to take account of potential nonlinearities in returns to scale of social influence. For example, a focal actor who experiences the churn of multiple social contacts might show super-additive effects (i.e., a combined effect that is larger than the sum of the individual effects) if one loss reinforces and legitimizes the other, or sub-additive effects (i.e., a combined effect that is smaller than the sum of the individual effects) if each additional effect provides less incremental information than the previous one. Quadratic effects allow us to account for such mechanisms to a certain extent.

6. Data collection

Our study is integrated into a larger body of research that uses electronically recorded communication patterns to approximate social relationships within a larger network. The substantial effort that has traditionally been associated with collecting social network data (Reingen & Kramar, 1986), combined with the inherent measurement bias of self-reported information on social relationships (Feld & Carter, 2002), has motivated researchers to identify alternative data collection techniques for social network analysis. Recently, the use of large-scale databases has gained increasing popularity. Examples of applications in the marketing field include the work of Hill et al. (2006) and Goldenberg et al. (2009) in the area of new product adoption, as well as that of Trusov et al. (2009), who compared the effectiveness of word-of-mouth and traditional advertising. Our analysis follows this stream of research and, similarly to Nitzan and Libai (2011) and Haenlein (2011), uses the call history database of a cellular service provider to reconstruct the social network of a set of customers.

Specifically, our dataset stems from a mobile phone operator in a major European country and was collected between August 2006 and April 2007. In that period, there were five major mobile phone operators in this market, four (including our company) with market shares between 22% and 26% and one with a market share of 4%. The market is among the top ten markets for mobile phone services in Europe and can be considered to be very mature, with a penetration rate in excess of 100%. As is the case for most mobile phone operators in Europe, the billing scheme that is applied by our company follows the “calling party pays” principle. This implies that the mobile phone subscriber does not pay for incoming calls and only pays for outgoing calls (Valletti & Houpis, 2005).

Our main data source was the customer database of the mobile phone operator, which includes basic information on each customer (e.g., acquisition date) and information on all of the incoming and outgoing calls that were made (i.e., calling party, calling time and calling duration). Our first step in collecting the data was to select a random sample of customers from the database. Because every customer

---

2 As has been discussed by Manski (1993) and Moffitt (2001), among others, the analysis of social contagion is subject to a series of potential biases that are introduced by simultaneity (i.e., person A’s actions affect person B’s actions and vice versa), correlated unobservable variables (i.e., omitted variables that affect the actions of A and B) and endogenous tie formation (i.e., the formation of ties that are based on common behavior). As shown by Lin and Wei (1989), Cox Proportional Hazards models with a non-parametric baseline hazard provide robust parameter estimates, even if the model is misspecified. This finding implies that the inclusion of a non-parametric baseline hazard prevents these problems to some extent, including, specifically, the bias that is introduced due to omitted correlated unobservable variables.

3 Mobile phone providers (including the one we collaborated with) usually sell a wide range of different service plans that represent contractual relationships of a fixed duration. Once a contract expires, the user has the option to subscribe to a new contract. Contract duration, which is the duration of the current contract, therefore differs from the total duration of the customer–company relationship (i.e., the time elapsed since acquisition).

4 Note that while we know when each customer has been acquired (i.e., how long she or he has been a customer of the mobile phone operator) and when the current contract the customer is subject to was established, we do not know how much time is left on the current contract. We return to this caveat when we discuss the limitations of our analysis.
can be uniquely identified by his or her mobile phone number, we generated a list of 1.25 million random numbers and matched them to the customer database. This process resulted in a random sample of 4163 customers (100%). We then removed 424 customers from our dataset who had joined the cellular service after August 1, 2006 (the beginning of the data collection timeframe) and for whom, therefore, only partial information was available, which resulted in 3739 customers (90%). Out of these 3739 customers, 3450 customers (83%) could be matched to a second database containing churn and revenue data, and of these, 3431 customers (82%) could be matched to a third database that contains demographic data. For these 3431 customers, we downloaded information about all of the incoming calls that were received and outgoing calls that were made (phone number and duration of call) over a two-month time period (August 1, 2006 to September 30, 2006, time period 1). This information was subsequently used to define the incoming and outgoing calling relationships and to determine tie strength according to Eqs. (1) and (2).

Our network, which represents the egocentric networks of the final sample of 3431 focal actors, includes 19,668 nodes that are linked through 25,799 calling relationships, or ties. On average, over the course of time period 1, each focal actor maintained 1.07 exclusively incoming calling relationships (corresponding to alters who call the actor but whose calls are not reciprocated), 1.01 exclusively outgoing calling relationships (corresponding to alterns who are called by an actor but who do not reciprocate the call) and 2.72 mutual calling relationships (see Table 2 for a full distribution of the number of calling relationships per focal actor). The total number of calling relationships that were maintained by each focal actor is hence 4.8 (1.07 + 1.01 + 2.72). While this finding might appear to be relatively low, it should be noted that our analysis only covers calling relationships in which both partners are customers of the focal cellular service provider (a necessary condition that ensures our ability to evaluate the churn behavior of both parties). Tie strength ranges from a minimum of 0% to a maximum of 100% for both incoming and outgoing relationships. The average tie strength in our sample was 7.1% for incoming relationships and 4.2% for outgoing relationships.

A total of 547 focal actors churned between November 1, 2006, and April 30, 2007 (time period 2), which corresponds to a churn rate of 15.9% (the overall churn rate of the company during this period was approximately 15.5%). Approximately 80% (447) of these churn events were voluntary, and 20% (100) were involuntary; the majority of the latter (95) were terminations that were due to bad debt (i.e., non-payment of outstanding bills). Among the 3431 focal actors, 1213 (35.4%) had at least one social contact who churned within the observation period. Out of those, 155 focal customers subsequently churned themselves, which represents a total of 28.3% of all of the 547 focal actors who churned. In terms of descriptive statistics, the average age was 36.2 years for voluntary churners, 33.5 years for involuntary churners, 38.9 years for non-churners and 38.4 years overall. The average contract duration was 1.4 years (504 days) for voluntary churners, 0.9 years (326 days) for involuntary churners, 1.4 years (500 days) for non-churners and 1.4 years (495 days) overall. The average duration of staying with the company (i.e., time elapsed since acquisition) was 2.3 years (858 days) for voluntary churners, 1.8 years (651 days) for involuntary churners, 4.2 years (1535 days) for non-churners and 3.9 years (1421 days) overall. A full list of descriptive sample statistics can be found in the Web Appendix.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Voluntary churn</th>
<th>Involuntary churn</th>
<th>No churn</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of incoming calling relationships</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>231</td>
<td>40</td>
<td>1341</td>
<td>1612</td>
</tr>
<tr>
<td>1</td>
<td>141</td>
<td>28</td>
<td>873</td>
<td>1042</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>13</td>
<td>348</td>
<td>399</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>7</td>
<td>164</td>
<td>192</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>6</td>
<td>70</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0</td>
<td>41</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9+</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>447</td>
<td>100</td>
<td>2884</td>
<td>3431</td>
</tr>
<tr>
<td><strong>Number of outgoing calling relationships</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>68</td>
<td>21</td>
<td>354</td>
<td>443</td>
</tr>
<tr>
<td>1</td>
<td>135</td>
<td>24</td>
<td>794</td>
<td>953</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>22</td>
<td>568</td>
<td>688</td>
</tr>
<tr>
<td>3</td>
<td>57</td>
<td>6</td>
<td>399</td>
<td>462</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>9</td>
<td>250</td>
<td>298</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>5</td>
<td>175</td>
<td>197</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>3</td>
<td>103</td>
<td>118</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>3</td>
<td>63</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td>9+</td>
<td>10</td>
<td>5</td>
<td>133</td>
<td>148</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>447</td>
<td>100</td>
<td>2884</td>
<td>3431</td>
</tr>
</tbody>
</table>

7. Estimation results

We estimated the model that is specified in Eqs. (1)–(2), (6)–(7) and (9)–(10) using the survival package (Version 2.35-4) within the R computing environment (Version 2.12.1). Specifically, we encapsulated the likelihood of the stratified Cox Proportional Hazards model in a constrained optimization algorithm to identify optimal values for $\alpha$ and $\beta$ ($\alpha, \beta > 0$) that result in maximum likelihood, given an estimation of $\gamma$ contingent on $\alpha$ and $\beta$. The optimization algorithm that was used is a limited-memory modification of the BFGS quasi-Newton method.

7.1. Model comparison

We specified and compared five different models (elaborated in Table 3). Model 1 (Model 2) incorporates no-churn recency and assumes that the two social impact measures influence the hazard of churn in a linear (quadratic) way; Model 3 is a baseline model that includes only the covariates without any consideration of social impact measures; Model 4 and Model 5 mirror Model 1 and Model 2 but allow for churn recency effects. Table 3 shows the fit of our final model (Model 5), as well as of the rival model specifications. Model fit was measured by the Bayesian information criterion (BIC) and by the Akaike information criterion (AIC). Two findings are of particular interest.

5 In total, we observed 3674 exclusively incoming calling relationships, 3475 exclusively outgoing calling relationships and 9325 mutual calling relationships. The total number of ties (25,799) was equal to the total number of incoming ties (3674 + 9325 = 12,999) plus the total number of outgoing ties (3475 + 9325 = 12,800).
First, models that take churn recency into account (Model 4 and Model 5), i.e., which weight churn events differently depending on how long ago they occurred, show better fit than models that do not (Models 1–2). Interestingly, when churn recency is not explicitly accounted for, a model without any social impact measures (Model 3) fits the data better than one that takes churn into account among incoming and outgoing calling relationships (Models 1–2). This surprising result can be explained as follows: As we will show later, only churn events that occur within a relatively small window of time are significantly related to focal actor churn. Considering all churn events to be independent of their recency therefore introduces a substantial amount of noise into the calculation of the social impact measures. Our results show that the inclusion of these noisy variables into the model deteriorates model fit instead of improving it.

Second, once churn recency is taken into account, the inclusion of churn among incoming and outgoing relationships does improve model fit. This finding implies that, in line with our basic hypothesis, social interactions affect customer churn decisions. The best-fitting model is Model 5, which includes the two social impact measures in quadratic form. This finding implies that the influence of social impact measures on focal actor churn is nonlinear. Model 5 has the lowest BIC and AIC of all of the models and, as indicated by an analysis of deviance, it represents a significant improvement compared to both Model 3 and Model 4.\(^6\)

7.2. Churn recency

The optimal values for the weighting function that result in maximum likelihood for the stratified Cox Proportional Hazards model are \( \alpha = -28.00 \) and \( \beta = -1.20 \). Fig. 1 illustrates the shape of the weighting function that corresponds to these two values. The churn events that occurred less than two weeks prior to \( t \) are accounted for nearly to their full extent in the calculation of \( s_{	ext{in},i}(t) \) and \( s_{	ext{out},i}(t) \) (\( w > 0.90 \)), while the churn events that happened more than five weeks prior to \( t \) have essentially no impact (\( w < 0.10 \)). Between these two extreme values, the weighting function follows a sigmoid shape, which is characteristic of the Gompertz function. From a managerial perspective, this finding implies that once a churn event has occurred, a company has a very small window in which it can attempt to prevent churn among the defecting customer’s social contacts. We will come back to this point in our discussion section.

\(^6\) The \( \chi^2 \) difference (\( -2 \times \) Difference in LL) between Model 3 and Model 4 is 27.26, and the \( \chi^2 \) difference between Model 4 and Model 5 is 20.00.

7.3. Parameter estimates—Control variables

Table 4 provides the detailed results for Model 5, i.e., parameter estimates, standard errors and \( p \)-values, as well as the parameter estimates for a baseline model without social effects (Model 3) for comparison. Looking at the demographic control variables, we see, on the one hand, that age is significantly negatively related to churn (\( -0.02 \)). This implies that customers become less likely to churn with increasing age, which is consistent with prior literature on the impact of age on repeat buying (Lambert-Pandraud & Laurent, 2010; Lambert-Pandraud, Laurent, & Lapersonne, 2005). On the other hand we observe that widowed customers are more likely to churn (3.41). Because these customers tend to be older (the mean age among widowed customers is 63.25 years, ranging between 51 and 83), this might be seen as a contradiction to the previous finding. However, because prior research has highlighted that elderly people are more vulnerable to fraudulent and unethical marketing practices (Yoon et al., 2005), it is conceivable that the relationship between age and loyalty might not be linear but an inverse U-shaped one. Our empirical findings support this notion.

With respect to socio-demographics, we see that, with one exception (Group 07), all of the variables that indicate socio-demographic group membership are significantly positively correlated with churn propensity. Group 12 has the highest coefficient (2.86) and corresponds to customers for whom socio-demographic group information is not available (“unclassified”). The second-highest coefficient (2.08) corresponds to Group 08, which represents pensioners with an active lifestyle and an above-average income. Again, this finding shows that customers with a relatively high age tend to be more likely to churn than other types of clients, which is consistent with our previous results.

Finally, regarding acquisition process characteristics, one can see that two out of twelve dummy variables for the acquisition channel are significant. Channel 07 corresponds to retail outlets that are directly maintained and owned by the cellular service provider itself. The negative coefficient (\( -0.65 \)) implies that customers that are acquired directly have a lower-than-average likelihood of churning. Channel 11 represents an internal channel that handles customers who have already threatened to leave a company in the past. The significant positive coefficient that is associated with this channel (1.75) shows that customers who have been retained using special means tend to be less loyal in the long run than other types of clients.

7.4. Parameter estimates—Social impact measures

Looking at the main variables of interest for our study, the impact of churn among incoming and outgoing calling relationships, two points are particularly noteworthy: First, while the two impact measures that are related to churn among outgoing calling relationships influence focal actor churn significantly (\( p \)-value \( \leq 0.00005 \)), the two impact measures that are related to churn among incoming calling relationships do not (\( p \)-value \( \geq 0.16 \)). This finding implies that, consistent with our theoretical expectations, the direction of the calling relationship matters because only churn among outgoing calling relationships influences the hazard of churn for the focal actor. The fact that we observed different results depending on relationship directionality supports the assumption that the significance of social impact measures is unlikely to be caused by confounds such as endogenous group formation, correlated unobservable variables and simultaneity (Manski, 1993; Moffitt, 2001) and provides an empirical indication for the presence of social contagion effects.

Second, churn among outgoing calling relationships influences the hazard of churn for the focal actor in a nonlinear way. For low levels of \( s_{	ext{out},i}(t) \), a focal customer’s hazard of defection increases with increasing numbers of defections among the social contacts with whom he or she has outgoing calling relationships. For example, a focal actor who
experiences the churn of an outgoing calling relationship with a recency-weighted tie strength of 1% (or, alternatively, of multiple outgoing calling relationships with a total tie strength that is equal to 1%) has a hazard rate that is approximately 1.6 times larger than that of a customer who was not exposed to churn among outgoing calling relationships. If, however, churn among outgoing calling relationships increases from 1% to 2%, this ratio increases from 1.6 to 2.5.7

To better interpret these results, it is interesting to contrast them to a model without social influence. The parameter estimates for such a model (Model 3) are shown in the first column of Table 4. As observed, a model without social effects results in very similar parameter estimates for all of the other variables than one that includes social effects. All of the coefficients have the same sign (with the exception of acquisition channel 3, a variable that is not significant in both models) and are of comparable orders of magnitude. Looking at the parameter estimates of Model 3 also helps to interpret the change in the hazard ratio that is associated with a certain social effect. For example, we see that the churn of an outgoing calling relationship with a recency-weighted tie strength of 2% (which is equal to 2.5, as discussed above) corresponds to a change in age of approximately 40 years.8 This finding shows that, at least in our sample, social effects have a considerable impact on the hazard of churn.

8 Robustness checks

Given the significant difference we observed in the impact of churn among incoming versus outgoing calling relationships, it is conceivable that other factors in addition to relationship directionality influence the importance of certain churn events. To test for such effects, we estimated three variations of Model 5 in which we classified incoming and outgoing calling relationships based on whether the two relationship partners (a) share the same service plan, (b) share the same acquisition channel or (c) share the same ZIP code. An analysis of deviance for the three alternative models on four degrees of freedom (df; each model includes four additional parameters) results in χ² values of 9.8 for service plan, 2.3 for acquisition channel and 6.5 for ZIP code.9 This shows that two of these three alternative models (those that classify relationships according to same or different acquisition channels or ZIP code) yield no significant improvement over Model 5. The improvement of the fit of the third model (classifying relationships according to same or different service plans) is significant with a p-value of 4.42%.

Table 5 shows the estimation results for the eight social impact measures (s_in, and s_out, for same and different service plans in both linear and quadratic forms) for a model in which relationships are distinguished according to service plan (Model 5.1). As observed, our substantive conclusions also hold in this model: churn among incoming calling relationships has no significant impact on focal actor churn, while churn among outgoing calling relationships impacts focal actor churn significantly.

We also tested a variation of Model 5, in which we differentiated social impact measures that were based on alter churn type (voluntary versus involuntary), which allowed us to test for differences regarding these two types of churn in the parameter estimates in addition to the unspecified baseline hazard function. This alternative specification did not improve model fit significantly (χ² difference of 5.8 on 4 df).

It is interesting to note that the time-dependent covariate “contract duration” is not significantly related to the hazard of churn within in Model 5 (p-value: 0.5238).10 This finding might seem surprising and could be considered as an indication of a confounding effect between contract duration and the social impact measures: Assume a focal customer who, once acquired, does not subscribe to a new contract and who experiences churn within her social network early in the company relationship. For this client, contract duration would increase over

---

7 The change in the hazard rate that is associated with a certain value of recency-weighted s_out, i can be determined as: \( \exp(0.4715 \times s_{\text{out}, i}) - 0.0087 \times s_{\text{out}, i}^2 \). It should be noted that the increase in the hazard of defection with increasing levels of churn among outgoing relationships is only true for low levels of recency-weighted s_out, i up to a maximum value of 27.0%. Beyond that point, the hazard of defection decreases with increasing levels of churn among outgoing relationships. That said, recency-weighted churn among outgoing calling relationships ranges from a minimum of 0% to a maximum of 100% in our analysis sample, with a mean of 1.05% over all observations and a mean of 3.91% for all cases in which recency-weighted s_out, i is larger than zero. The results for high levels of recency-weighted churn among outgoing calling relationships therefore need to be interpreted with caution.

8 Calculated as \( \ln(2.5) / 0.0223 \).

9 The critical χ² value at 4 df is 9.49.

10 In fact, contract duration is not a significant predictor in any of the models that are shown in Table 3 (p-values equal to or above 0.30).
improvement in model fit was a model with a threshold at 205 days ($\chi^2$ difference compared to Model 3 of 10.15 on 1 df, p-value: 0.0014). Within this model, contract duration influences the hazard of churn positively (coefficient: 0.0062) when it was below the threshold of 205 days (p-value: 0.0023) and was not significant otherwise. The fact that the result regarding contract duration was identical in models with and without social impact measures makes the presence of any confounding effects unlikely.

Finally, we estimated a variation of Model 5 using a different specification of the contract duration variable. In the original specification of Model 5, we operationalized contract duration as the duration of the current contract. This measure was potentially subject to substantial, discrete shifts because contract duration increases steadily until a new contract is signed, and the measure falls back to zero. To test the extent to which these discrete shifts create problems in the parameter estimation process, we estimated an alternative model (Model 5.3) in which we defined contract duration as the total duration of the customer–company relationship (i.e., time elapsed since acquisition). As observed in Table 5, this again did not affect our substantive conclusions regarding $\delta_{\text{contract}}$ and $\delta_{\text{client relationship}}$.

9. Discussion

9.1. Theoretical implications

In terms of contribution to theory, our results provide three important theoretical implications: First, consistent with the work of Nitzan and Libai (2011), we find an empirical indication for social interactions in customer churn decisions. Churn of customers with whom a focal actor maintains outgoing calling relationships has a significant and substantial impact on the hazard of churn of the focal actor. The fact that we used a model that takes account of network directionality and includes controls for homophily, correlated unobservable variables, and the characteristics of the acquisition process and contract reduce the likelihood that our findings are caused by spurious effects and provides an indication for the presence of social contagion rather than a simple correlation in observed behavior. Our results therefore confirm that social network effects, which have previously mainly been discussed in the context of customer acquisition, are also present in the case of customer retention and hence remain important even after customers have been acquired. Specifically, they imply that

---

11 We thank an anonymous reviewer for having pointed this out.
individual-level retention rates and customer lifetime have a certain degree of positive network autocorrelation.

Second, our work provides insight into the dynamics that underlie social influence in customer retention by showing that a focal actor’s likelihood of churn is differentially influenced by churn among his or her social contacts and that this influence is dependent on relationship directionality and churn recency. Specifically, we show that only churn events among outgoing calling relationships are related to focal actor churn and that these events affect an actor’s likelihood of churning at a given time only if they occurred within a short period (in our sample, less than five weeks) prior to that time. If churn recency is not explicitly accounted for, then churn among a focal actor’s social relationships does not have a significant impact on the likelihood that the focal actor will churn. Furthermore, the influence of churn among social relationships on the hazard of churn for a focal actor is nonlinear in the sense that the hazard of defection increases quadratically with increasing numbers of recent churn events among outgoing calling relationships.

Combined, these results allow us to speculate about the drivers behind social influence processes, which are very often difficult or even impossible to distinguish empirically (DiMaggio & Powell, 1983; Van den Bulte & Lilien, 2001). For example, one possible interpretation of our findings is that customers simply try to minimize cost and therefore base their retention decisions on churn among their social relationships. Because within our example only outgoing calling relationships represent a cost for a focal actor (due to the “calling party pays” principle), high levels of churn among these relationships provide an incentive for defection to minimize the share of expensive out-of-network calls. However, while such an interpretation can explain the importance of churn among outgoing versus incoming calling relationships, it is not consistent with the importance of churn recency that we observed in our sample. Given that the accumulated cost for out-of-network calls increases with the amount of time that elapsed since the churn of a social contact, we would expect that churn events among outgoing calling relationships that happened longer ago should have the same, if not a higher, impact on focal actor churn than recent events. Instead, we observed the opposite result in our sample. This decay of social influence is consistent with the previous research on the exponential decay of word-of-mouth effects over time (Strang & Tuma, 1993; Trusov et al., 2009). This finding implies that our results are more consistent with an interpretation of social contagion than of simple cost minimization.

Third, our finding that relationship directionality plays an important role in determining the likelihood that a focal actor will churn following the churn of his or her peers also has implications for other research contexts that make use of electronically recorded data on social relationships. Researchers increasingly use data from email networks (Kossinets & Watts, 2006), internet communities (Grabowski, 2007), social networking sites (Trusov et al., 2009) and call detail records (Eagle, Pentland, & Lazer, 2009; Onnela et al., 2007; Palla, Barabasi, & Vicsek, 2007) to reconstruct social networks. Some of these studies have resulted in surprising and sometimes counterintuitive findings, for example on the importance of strong versus weak ties (Onnela et al., 2007) or the degree of overlap between observational data from mobile phones and self-reported survey data (Eagle et al., 2009). The same is true for certain simulation studies that have relied on undirected networks and provided results that contradict commonly held beliefs (Watts & Dodds, 2007). Given that ties of different directions might have different importance for social processes, it is conceivable that these counterintuitive results may in part be driven by the use of undirected rather than directed networks. We therefore encourage researchers to take tie directionality into account in both empirical and simulation works, especially when relying on electronic tie data. If possible, researchers should prefer directed data over undirected data.

9.2. Managerial implications

From a managerial viewpoint, our analysis indicates that monitoring churn within the social network of a customer and specifically among outgoing relationships can provide useful information in the context of churn prediction and proactive churn management. As shown by Neslin et al. (2006), even a small improvement in the accuracy of churn prediction models can be associated with substantial financial gains for companies. Especially in the telecommunications industry, where customer churn has been shown to be a major cost item and where information on social relationships can be obtained easily through the analysis of call detail records, calculating the impact measures of social network churn is likely to be particularly beneficial.

That said, the high importance of churn recency that we observed in our sample implies that, following a customer’s churn, companies only have a very small window in which they can attempt to prevent subsequent churn among the defective customer’s social contacts. Therefore, in many situations it might not be wise to wait until churn among social contacts actually occurs because at that point proactive churn prevention techniques might no longer be feasible. Instead, it seems more advisable to include the positive autocorrelation among individual-level lifetimes and churn propensities that we observe in our analysis within the churn prediction model itself to allow for more accurate predictions. Information on the estimated churn propensity of one customer could constitute additional input towards estimating the churn propensity of another client. Linear network autocorrelation models have previously been used in social network analysis to account for and measure social contagion effects (e.g., Leenders, 2002). Recent advances in spatial modeling now also allow the inclusion of network autocorrelation in logit models, which are frequently used to predict churn propensities (Carl & Kuehn, 2007; Dow, 2008).

9.3. Methodological contribution

In addition to these theoretical and managerial implications, our work also provides a methodological contribution by proposing an approach to incorporate continuous recency effects into Cox Proportional Hazards models. The usual approach to account for recency in nonparametric survival models has been to work with lagged variables as was done, for example, by Nitzan and Libai (2011). This technique, however, only tests for differences at discrete time points and does not allow a continuous recency effect to be included in the model estimation. Our approach, which uses a continuous weight function to weight churn events according to recency, overcomes this limitation.

9.4. Future research

With respect to areas of future research, we see three questions of particular importance: First, in some calling relationships, for example among spouses or other family members, the relationship partners might actively try to balance economic costs between each other by offering to terminate a call and to call the other person back. Such actions, if they occur, would represent a distortion of our measure of relationship directionality, especially in the case of mutual calling relationships. One way to assess the extent of bias that this distortion creates would be to test the robustness of our results when excluding calling relationships that belong to the same family or household. Another option would be to consider other measures (e.g., qualitative assessments) that characterize the type of relationship in addition to tie strength and relationship directionality. Additionally, as highlighted above, while we know the acquisition date and contract duration of each customer, we

12 We thank an anonymous reviewer for having provided this explanation.
do not know how much time is left on the current contract. It is therefore conceivable that this missing information could be confounded with the social impact measures that we include in our analysis. If possible, future studies should correct for this potential bias by considering the remaining contract duration explicitly in the model. In sum, these modifications would increase confidence in our results and also allow further insight to be provided into the relative importance of certain types of relationships in driving focal actor churn decisions.

Second, future studies could combine our finding of social network effects in individual-level churn behavior with recent evidence on social network effects in customer-level revenue (Hanlein, 2011), to provide insight into the extent of network autocorrelation in customer lifetime value. Customer lifetime value is a measure of the cumulative discounted profit that a customer generates over his or her lifetime. Given that individual-level churn behavior, customer lifetime and individual-level revenue are all subject to positive network autocorrelation, it can be expected that similar autocorrelation will be observed – possibly to an even greater extent – in customer lifetime value. Assessing the extent of this autocorrelation would provide valuable insights for customer relationship management and for managing the customer base as a whole versus on an individual basis. Traditional approaches to calculating customer lifetime value (e.g., Berger & Nasri, 1998; Fader, Hardie, & Lee, 2005) or agent-based models such as stochastic cellular automata (Goldenberg, Libai, & Muller, 2001) could be applied in this context.

Third, in recent years, research on social influence in customer acquisition has started to move from providing evidence that social contagion exists to giving explanations on why it occurs (Iyengar, Van den Bulte, & Choi, 2012). The literature has discussed several causal mechanisms of social influence, including information transfer, normative pressures, competitive concerns and performance network effects (Van den Bulte & Lilien, 2001). Applied to social interactions in customer churn decisions, this finding implies that different causal mechanisms might operate in parallel and explain the findings we observed. For example, following the churn of a social contact, a focal actor might be more likely to churn because she or he becomes aware of a more attractive offer (information transfer), becomes concerned that remaining with the old provider may result in status disadvantages (normative pressures), or seeks to reduce the costs that are incurred as a result of the differential pricing of in-network and out-of-network calls (performance network effects). A more detailed investigation of the mechanisms through which social contagion occurs will provide insights into the importance of each of these drivers, which could then be used for proactive churn management actions.

Appendix A. Weighting function

The Gompertz function \( f(x) \) with parameters \( a, b \) and \( c (b, c < 0) \) is defined as:

\[
f(x) = ae^{be^{cx}}. \tag{A1}
\]

Based on this definition of \( f(x) \), we define a new function \( g(x) \) as:

\[
g(x) = a - f(x) = a - ae^{be^{cx}} \text{ with } a = \frac{1}{1-e^{b}}. \tag{A2}
\]

which is equivalent to:

\[
g(x) = \frac{1}{1-e^{b}} \left( 1 - e^{be^{cx}} \right). \tag{A3}
\]

It can easily be observed that \( g(x) \) fulfills the following three conditions:

- \( g(0) = \frac{1}{1-e^{b}} \left( 1 - e^{b} \right) = 1 \)
- \( \lim_{x \to -\infty} g(x) = \frac{1}{1-e^{b}} = \frac{1}{1-\epsilon^{b}} = 0 \)
- \( g(x) \geq g(x + c) \text{ for all } x \)

This implies that \( g(x) \) can be used as a weighting function in defining \( s_{\text{src}}(t) \) and \( s_{\text{tar}}(t) \).

The function \( g(x) \) allows for a large variety of different shapes, depending on the choice of \( b \) and \( c \). Specifically, we see that:

- \( \lim_{x \to 0} g(x) = \frac{1}{1-e^{b}} \left( 1 - e^{b} \right) = 1 \)
- \( \lim_{b \to 0} g(x) = e^{cx} \)

The limiting cases of \( g(x) \) are therefore equal weighting of all churn events (for \( c \to 0 \)) and exponential decay (for \( b \to 0 \)).

Appendix B. Calculation of social impact measures

The use of time-dependent covariates (i.e., contract duration and churn among incoming or outgoing calling relationships) made it necessary to transform the input data before using it for model estimation. Table B.1 illustrates the process of this data transformation. The upper panel shows an input dataset for an actor who maintains four different calling relationships with three alters. The lower panel (the analysis dataset) shows the transformed version of this input data, which serves as the input to estimate the stratified Cox Proportional Hazards model.

The first line in the analysis dataset covers the period between the actor acquisition date (01.01.2005) and the churn date of the first alter (Alter ID 2, 15.11.2006). Because no churn occurred among any of the alters during this period, the impact measures for incoming and outgoing relationships are zero. Given that the current contract of the actor started on 01.07.2006, the contract duration at the churn date of the first alter is 137 days.

The second line covers the period between the churn of Alter ID 2 (15.11.2006) and the churn of the Alter ID 3 (15.12.2006). The focal actor maintained an outgoing calling relationship with Alter ID 2, with a tie strength that was equal to 0.05. The time difference between the churn date of Alter ID 3 (15.12.2006) and the churn date of Alter ID 2 (15.11.2006) is 30 days. Therefore, the outgoing impact measure for this period is 0.05 \( \times w(30) \), where \( w(30) \) is the value of the weighting function for 30 days. The incoming impact measure remains zero. The focal actor’s contract duration at the churn date of Alter ID 3 is 167 days.

The third line covers the period between the churn of Alter ID 3 (15.12.2006) and the focal actor’s churn date (01.01.2007). The focal actor maintained a mutual calling relationship with Alter ID 3. The tie strength of the incoming calling component of this relationship is equal to 0.10, and the tie strength of the outgoing calling component is equal to 0.15. The time difference between the actor’s churn date (01.01.2007) and the churn of Alter ID 2 (15.11.2006) is 47 days, while the time difference between the actor’s churn date and the churn date of Alter ID 3 (15.12.2006) is 17 days. The outgoing impact measure for this period is therefore 0.05 \( \times w(47) + 0.15 \times w(17) \), and the incoming impact measure is 0.10 \( \times w(17) \). Note that because Alter ID 3 is a mutual calling relationship, both social impact measures \( s_{\text{src}} \) and \( s_{\text{tar}} \) increase. The increase in each variable depends on the proportion of the total incoming or outgoing communication volume that Alter ID 3 accounts (i.e., tie strength). The contract duration at the actor’s churn date is 184 days.

To summarize, the transformed dataset that is used to estimate the stratified Cox Proportional Hazards model includes \( n + 1 \) data points for each customer, where \( n \) is the number of alters who have churned prior to the actor’s churn date or the end of the observation period. Note that within our example, the churn of the last alter (Alter ID 4) is irrelevant because it occurred after the actor’s churn date.
Appendix C. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jiresmar.2013.03.003.

References


**Pricing in the international takeoff of new products**

Deepa Chandrasekaran a, Joep W.C. Arts b, Gerard J. Tellis c,*, Ruud T. Frambach b

a University of Texas at Austin, United States
b VU University Amsterdam, The Netherlands
c Marshall School of Business, University of Southern California, United States

**ABSTRACT**

This study focuses on the effect of two dimensions of price (relative price and price volatility) on the international takeoff of new products. The study examines these drivers of takeoff using a novel data set of bimonthly observations of 7 new consumer electronic products in 8 countries. The empirical analysis reveals that both relative price and price volatility significantly impact the hazard of takeoff. However, although the effect of relative price is stable across contexts, the effects of price volatility are moderated by wealth, culture, and contagion. The use of temporally disaggregate data at the bimonthly level allows for the identification of the effect of price volatility and enables a more precise identification of takeoff than that achievable with annual data.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Takeoff refers to the first dramatic increase in the consumer adoption of new products, marking a transition from the introduction stage to the growth stage of the product life cycle (Golder & Tellis, 1997). Recent research has focused on the identification and patterns of takeoff (Agarwal & Bayus, 2002; Golder & Tellis, 1997), as well as on gaining an understanding of the drivers of takeoff across countries (Chandrasekaran & Tellis, 2007, 2008; Muller, Peres, & Mahajan, 2009; Tellis, Stremersch, & Yin, 2003; Tellis, 2013; Van Everdingen, Fok, & Stremersch, 2009). With respect to this literature, the current study makes two key contributions. First, the study highlights the role of two key dimensions of price that may affect the takeoff of new products in an international context: relative price and price volatility. Second, the study examines the role of country-level macro and contagion factors that may moderate the role of the pricing dimensions on takeoff.

The extant literature has considered the impact of relative price, measured as the current price relative to the initial price of the product, on new product takeoff in national contexts. The idea is that consumers respond to decreases in the prices of new products, particularly during the product’s initial years, and this hastens takeoff (e.g., Golder & Tellis, 1997; Wei & Xiao, 2012). Although this research has generated insights into the effect of pricing on takeoff, it is not clear whether these effects hold in an international context and, if so, in which ways. We therefore examine the effect of relative price, accounting for possible influences across countries. The general marketing literature also suggests that consumers’ purchase decisions are influenced by short-term pricing volatility, which may lead to a deviation in actual price from an expected price (e.g., Winer, 1986). That is, although a long-run price decrease may be expected, the industry witnesses much short-term price volatility, marked by either price increases or decreases. Fig. 1a, b, and c illustrate the extent of volatility in the product pricing history for several brands in various categories of new consumer electronics. Indeed, the industry has recently experienced the rise of several websites (such as decide.com, nextag.com, and shopobot.com) devoted to facilitate consumer purchase decisions (buy or wait; compare) in the face of uncertainty resulting from such pricing (e.g., Geron, 2011; Needelman, 2011). However, the effect of short-term price volatility on product takeoff has not been examined.

The literature on international takeoff examines the specific impact of cross-national factors on the hazard of takeoff, largely ignoring the role of strategic factors such as price (see Table 1). We not only examine the role of price in conjunction with broader macro factors but also examine the moderating role of important variables such as culture and wealth on the effect of pricing on takeoff. For example, some environments, such as those characterized by greater wealth, may be more receptive to relatively higher prices or more tolerant of higher price volatility than other environments, such as those characterized by lower levels of wealth. Motivated by these issues, the current study attempts to answer the following research questions:

1. What role do relative price and price volatility play in the international takeoff of new products?
2. What factors moderate the role of these two variables in the context of international takeoff?

Our contributions are facilitated by the use of disaggregate data (at the bimonthly level). The study uses a unique data set of bimonthly observations of 7 new consumer electronics products in 8 countries. With the exception of a few studies (Goldenberg, Lowengart, & Shapiro, 2009; Honisch, Pittnauer, & Stauffer, 2008), much of the academic research on new product growth uses annual data because such data are widely available (e.g., Agarwal & Bayus, 2002; Markovitch & Golder, 2008). However, because of the rapid adoption of categories of new consumer electronics in recent times, managers must model new product takeoff using more immediately accessible monthly, bimonthly, or quarterly data. For instance, Apple recently launched the iPad, stimulating the takeoff of media tablets.1 Media tablets sold more than 3 million units in the first few months after the introduction of the iPad. Such rapid takeoffs cannot be easily pinpointed using aggregate annual data. The use of temporally disaggregate data has several advantages for this study. First, fluctuations observed in more granular data do not necessarily constitute noise but may provide valuable information (Goldenberg et al., 2009). In this study, we are able to observe and measure price volatility at an inter-temporal level and identify its role with regard to takeoff. For instance, Fig. 2a and b provide the example of the plasma TV in the United Kingdom, comparing annual price data with price data at the bimonthly level. The annual graph indicates that the overall price trend is a decrease after the introductory price. However, the bimonthly graph indicates the extent of intertemporal price volatility with regard to this trend. Second, temporally disaggregate data allow for more efficient estimates (Clarke, 1976; Judge, Griffiths, Carter Hill, Lütkepol, & Lee, 1985; Tellis & Franses, 2006). In this study, the use of such data enables a fine-grained identification of takeoff.

The remainder of the paper is organized as follows. We first discuss the theory underlying our model and formulate hypotheses regarding the major contributions of this paper. We then discuss the methods used for the empirical research, including the data, the measures, and the model. Next, we present the findings of this study, along with several robustness checks. We conclude with a discussion of the study’s findings and implications.

2. Theoretical framework

Based on prior studies, this section develops a theoretical framework for the drivers of takeoff (Fig. 3).

The literature suggests that four latent factors underlie most explanations for the faster takeoff of certain products: A. within-country macro effects, B. market effects, C. contagion effects, and D. cross-country effects (e.g., Agarwal & Bayus, 2002; Chandrasekaran & Tellis, 2008; Golder & Tellis, 1997; Peres, Muller, & Mahajan, 2010; Tellis et al., 2003; Van Everdingen et al., 2009).

In terms of country-level factors, the key discussion centers on the role of economics, which is measured predominantly in terms of national wealth, and culture, which is measured predominantly in terms of uncertainty avoidance (Chandrasekaran & Tellis, 2008; Golder & Tellis, 1997; Tellis et al., 2003; Van Everdingen et al., 2009). In terms of the market dimension, the takeoff literature suggests that two factors are important: relative price and innovative activity (Agarwal & Bayus, 2002; Bayus, Kang, & Agarwal, 2007; Chandrasekaran & Tellis, 2008; Golder & Tellis, 1997; Tellis et al., 2003). In terms of contagion effects, a direct examination of the role of intra-country and interpersonal interactions on takeoff, either via word-of-mouth (W-O-M) or signals, has not been performed. In terms of cross-country effects, much of the takeoff literature has focused on the effect of prior foreign takeoffs on takeoff in focal countries (Chandrasekaran & Tellis, 2008; Tellis, Stremersch, & Yin 2003; Van Everdingen et al., 2009).

We contribute to the literature by examining a contingency framework for the effect of pricing on takeoff while controlling for other major drivers. At the core of our framework is the direct impact of pricing in terms of relative price and price volatility on product takeoff. Although the impact of relative price on takeoff has been examined, albeit not in an international context, the temporally

---


---

Fig. 1. a. Price History of a Digital SLR Camera Brand (from NexTag, showing Maximum, Median and Minimum prices). b. Price history of a GPS brand (from NexTag, showing Maximum, Median and Minimum prices). c. Expectations of a Short-run Price Increase for a LED tv brand (from Decide).
disaggregate nature of our data permits us to examine the importance of a second important dimension of pricing, i.e., price volatility. The following sections propose hypotheses regarding the main effect of two price dimensions and the moderating influence of country and contagion factors in the influence on takeoff. In addition, we discuss relevant control variables.

2.1. Price

Pricing is a critical variable determining the behavior of both the firm and the consumer (Tellis, 1986). For the firm, pricing determines an innovation's profit potential (e.g., Simon, 1992). For the consumer, pricing represents the cost to be paid for the benefits he/she receives and functions as signal of quality (e.g., Zeithaml, 1988). We consider two dimensions of pricing. First, we study the impact of the relative price, which we define as the current price relative to the introductory price. Second, we consider price volatility, which we define as the influence of short-run deviations from the expected trend in price. Below, we formulate hypotheses on both effects.

2.1.1. Relative price

Prior research on new product pricing suggests that relative price plays an important role in stimulating the sales growth of new products. The premise is that consumers do not respond directly to an absolute price, but rather, relative to the reference price (Rajendran & Tellis, 1994; Thaler, 1985). According to prospect theory (Kahnemann & Tversky, 1979) or accountancy (Thaler, 1985), if a consumer encounters a brand (or a product) at a price that is lower than the reference price, the price is perceived as a gain, whereas a price that is higher than the reference price is perceived as a loss. Hence, consumers consider not the absolute price but gains or losses relative to a reference price. In new product research, a price is often considered relative to introductory price. As the price of a product decreases over time, it leads to an increase in the adoption of new products. This trend may occur for several reasons.
First, Bagwell and Riordan (1991) argue that for new products, initial high prices signify quality and subsequent decreasing prices reflect the diffusion of product information and the adaptation of the price signal to consumer learning about the quality of the product. As consumers gain experience with the product and information about its quality diffuses, firms may be able to efficiently signal quality at lower prices (Bagwell & Riordan, 1991).

Second, consumer heterogeneity in price sensitivity drives takeoff. As the price of the new product decreases over time, the product becomes attractive to more price-sensitive customers and, as a critical mass of those customers adopt, takeoff occurs (Golder & Tellis, 1997). Indeed, economic research has demonstrated that because the purchase of new products is risky and their demand is price sensitive, the optimal pricing strategy for new products is often considered to be the monotonic lowering of price (Krishnan, Bass, & Jain, 1999). Typically, during the introduction stage of a new product, innovators and early adopters are attracted to the new product. These consumers are more likely than later adopters to favorably evaluate the utility/price tradeoff (Kalyanaraman & Winer, 1995), thus driving sales during the early stage of the product life cycle, even at high prices. In subsequent periods of the early life cycle, firms may strategically lower price to price discriminate among successively more price-sensitive segments (although as production costs decrease, prices may also decrease in response to the decreasing costs (Wei & Xiao, 2012).

Consistent with the above, meta-analytic studies on the effect of price on new product sales reveal that price elasticity is stronger during the introduction/growth stage than after takeoff, during the mature/decline stage (Bijmolt, Van Heerde, & Pieters, 2005). The reasoning behind this tendency suggests the following hypothesis:

**H1.** A lower relative price increases the hazard of takeoff for a new product.

### 2.1.2. Inter-temporal pricing volatility

New product pricing is characterized by both upward and downward movements, as well as significant changes within a short period of time. Agarwal and Bayus (2002) suggest that crucial R&D expenditures during the early years of market evolution may increase costs and translate into innovations that consumers value. Accordingly, firms may choose to recover R&D costs or exploit superior valuations by temporarily increasing prices.

![Fig. 2. a. Overall price decline visible using end-of-year data points for plasma TV in UK. b. Inter-temporal price volatility visible in bimonthly data for plasma TV in UK.](image-url)
How may short-term price volatility affect consumer behavior? Alba, Mela, Shimp, and Urbany (1999) show that consumers find it easier to estimate average prices in the case of consumer goods, where price differences tend to be small. However, price changes tend to be far greater for new consumer durables than for fast moving consumer goods. So, consumers experience greater difficulty in estimating an average price for new consumer durables (Marn, Roegner, & Zawada, 2003). Figs. 1a, b, and c and 2b illustrate price volatility situations that are typical with regard to the pricing of consumer electronics.

The higher the price instability, the less likely consumers are to reliably estimate a new product's price. Lacking the ability to form realistic expectations, consumers may postpone purchase of the innovation. In this vein, Winer (1985) posits that price volatility and unanticipated inflation or deflation may have a negative effect on demand. In the case of new products, consumers have multiple alternatives, including postponing their purchase or never buying the new product (Jacobson & Obermiller, 1990). Postponing adoption may be particularly conspicuous for new products that have not yet taken off because their success prior to takeoff is unclear to both firms and consumers. If too many consumers wait, sales of the new product remain low and the hazard of takeoff decreases. The above reasoning suggests the following:

**H2.** Higher inter-temporal pricing volatility decreases the hazard of takeoff for a new product.

2.2. Country factors and pricing

Prior research emphasizes two important country variables, national wealth and uncertainty avoidance, that influence product takeoff. We control for the primary effects of these variables in the regression analysis. In this section, we focus on how these variables may moderate the influence of pricing on takeoff.

2.2.1. National wealth and pricing

National wealth relates to consumers’ average spending power in a particular country (Talukdar, Sudhir, & Ainslie, 2002). The literature on international takeoff has supplied mixed evidence with regard to the role of national wealth in stimulating takeoff. Wealth was not found to have a significant effect on takeoff when a cluster of European countries were considered (Tellis et al., 2003) but was found to be a significant driver of takeoff when more heterogeneous sets of countries were considered (Chandrasekaran & Tellis, 2008; Van Everdingen et al., 2009).

How would national wealth influence the role of pricing in stimulating takeoff? We assume that on average, consumers in wealthier countries possess more disposable income than consumers in less wealthy countries. People are more price-sensitive when they have lower incomes and tighter budgets and less price-sensitive when they have higher incomes and greater spending power. Hence, wealthier consumers may be better able to afford new products whose performance is uncertain. Hence, we posit the following:

**H3.** The effect of a lower relative price on the hazard of takeoff is weaker in more wealthy countries than it is in less wealthy countries.

2.2.2. Uncertainty avoidance and pricing

Uncertainty avoidance refers to the extent to which the members of a culture feel threatened by uncertain or unknown situations.
Uncertainty avoidance has been considered the most important cultural factor affecting consumer innovativeness and product takeoff (Steenkamp, Hofstede, & Wedel, 1999; Van Everdingen et al., 2009). Prior research argues that consumers from countries with high uncertainty avoidance are often likely to buy products when uncertainty surrounding the new product's performance makes it difficult to evaluate price–benefit trade-off. Hence, consumers in high uncertainty avoidance contexts may be more motivated than consumers in low uncertainty avoidance contexts to wait until the relative price is substantially lower than the launch price. Hence, we posit the following:

**H5.** The effect of a lower relative price on the hazard of takeoff is stronger in countries with high uncertainty avoidance than in countries with low uncertainty avoidance.

Furthermore, in cultures with high uncertainty avoidance, consumers will focus on reducing risk and avoiding ambiguous situations. Hence, during times of higher price volatility, consumers may be more likely to adopt a “wait-and-see approach” to the purchase of consumer electronics. This reasoning leads to the formulation of the following hypothesis:

**H6.** The effect of price volatility on the hazard of takeoff is stronger in countries with high uncertainty avoidance than in countries with low uncertainty avoidance.

### 2.3. Contagion and pricing

The vast marketing literature on product diffusion has examined the influence of inter-personal word-of-mouth communication (variously termed as imitation and internal influence) on sales (e.g., Bass, 1969; Mahajan, Muller, & Bass, 1990). The recent literature has addressed the influence of additional phenomena such as social signals or information cascades within the scope of consumer interactions that may impact product diffusion (Peres et al., 2010). We use the term contagion synonymously with consumer interactions, which may refer to word-of-mouth communication, signals, or imitation, as described above. Below, we examine the moderating role of contagion with regard to the two dimensions of pricing.

#### 2.3.1. Contagion effects and relative price

One of the crucial influences driving consumers to purchase new products or services involves inter-personal communications or observations of the choices of others (e.g., Bass, 1969; Chevalier & Mayzlin, 2006; Ganesh & Kumar, 1996; Golder & Tellis, 2004; Mahajan et al., 1990; Peres et al., 2010; Putsis, Balasubramanian, Kaplan, & Sen, 1997; Tirunillai & Tellis, 2012; Van den Bulte & Stremersch, 2004). We use the term contagion to describe intra-country interactions among consumers.

How may contagion effects influence the role of relative price on the takeoff of new products? Anecdotal evidence suggests that when much conversation is generated about a product or if consumers are rushing to buy a product, some consumers may be motivated to buy the product immediately, irrespective of its price (Emigh, 2010; Heater, 2008). In fact, the literature on information cascades suggests that people converge in adopting a behavior with increasing momentum and declining individual evaluation because of their tendency to derive information from the behavior of prior adopters (Bikhchandani, Hirshleifer, & Welch, 1992; Golder & Tellis, 2004). For instance, Duan, Gu, and Whinston (2009) find that in the software adoption context, information cascades may lead to the adoption of inferior products by online users. Similarly, consumers may be more susceptible to higher-priced gadgets associated with higher levels of buzz, conversation and visible demand. Hence, we hypothesize the following:

**H7.** The effect of a lower relative price on the hazard of takeoff is weaker for a product with a high contagion effect than for a product with a low contagion effect.

#### 2.3.2. Contagion effects and price volatility

We propose that there is a negative interaction effect between inter-temporal price volatility and contagion effects. When inter-temporal price volatility is high, the perceived riskiness of buying the product immediately, as opposed to waiting, may be enhanced. However, a strong demand for the product or enhanced buzz/conversations may alleviate some of the concerns that may arise from buying the product immediately as opposed to postponing the purchase decision. Hence, we hypothesize the following:

**H8.** The effect of price volatility on the hazard of takeoff is weaker for a product with a high contagion effect than for a product with a low contagion effect.

### 2.4. Control variables

We control for the effects of 3 other important factors that are likely to influence takeoff. These variables include foreign takeoffs, innovative activity, and seasonality.

Prior research has determined that foreign takeoffs may accelerate takeoffs in a focal country (Chandrasekaran & Tellis, 2008; Tellis et al., 2003; Van Everdingen et al., 2009). We therefore control for the influence of foreign takeoffs in our model.

Earlier research has examined the main effect of innovative activity. Agarwal and Bayus (2002) suggest that the sales of new products may be initially low because the first commercialized forms of new products are relatively primitive. Firm entry may subsequently stimulate innovative activity, which increases the appeal of new products, widens their market, and stimulates takeoff (Bayus et al., 2007). Hence, we control for the effects of innovative activity to isolate the influence of pricing on takeoff.

Most products display periodic patterns throughout the year, a process referred to as seasonality (Radas & Shugan, 1998). With respect to consumer goods, seasonality is often caused by holidays (Christmas and New Year and spring and summer breaks) and changes in the weather, such as the start of a new season (Hylleberg, 1992; Miron, 1996). Many firms adjust their strategies to suit seasonal sales patterns (Axarloglou, 2003; Radas & Shugan, 1998). The momentum gained from seasonal promotion strategies may boost new product sales, increasing the likelihood of takeoff during peak seasons. We therefore control for seasonality in our model.

### 3. Method

This section describes the data, measures, and model used in this study.

#### 3.1. Data

We obtain data at the bimonthly level (2 consecutive months) from 1999 to 2005. The data cover the sale and prices of 7 new consumer durable products in 8 nations. Although limited by the nature of the disaggregate level of data, the number of unique product–country combinations in our data set is still higher than that in several prior studies examining the diffusion of new products (e.g., Agarwal & Bayus, 2002; Golder & Tellis, 1997, 2004; Helsen, Jedidi, & DeSarbo, 1993; Takada & Jain, 1991). The products include important innovations in the electronics industry introduced in the last 10 years and constitute innovations that have not been studied in the context of product diffusion. The
products include the Digital Video Recorder (a device that records video in a digital format to a disk drive or other memory medium within a device), the DVD Recorder (also known as the DVDVR, an optical disk recorder that records video onto blank writeable DVD media), the Surround Sound System (home theater), the LCD TV, the Plasma TV, the MP3 Hard Disk Player, and the MP3 Flash Player. Data are available for the United States and 7 European countries: France, Germany, Italy, the Netherlands, Spain, Sweden, and the United Kingdom.

We obtained the sales and price data from GFK International (www.gfk.com). In total, we obtained 1296 observations with regard to bimonthly sales and prices. GFK International collects the data based on actual sales at the retailer. We consider historical sales data from the time at which these products were commercialized. Countries are matched using time-origin for new product commercialization, in accordance with Dekimpe, Parker, and Sarvary (1998). The use of the term commercialization allows for the recognition that market research companies and databases include a product’s sales or penetration only when it has achieved a low level of sales or penetration. We determine the earliest available commercialization date through an examination of our bimonthly data and estimate the validity of these start dates through an assessment of penetration rates and a comparison with annual data sources, as performed in prior takeoff research. After our bimonthly data’s identification of the earliest start date, we begin our time count from the first data point. In 11 instances, annual sales and price data indicate that commercialization occurred 1 to 6 months earlier than the first available bimonthly data point and 10 instances in which the earliest commercialization occurred more than 6 months prior to our first bimonthly data point. For such cases, we consider the commercialization time to be the Dec-Jan period of the earliest available annual data point. In all of these cases, penetration at the first available data point is very low, with a mean of .05%.

We also collect information on different country characteristics from syndicated sources (Euromonitor Global Marketing Information Database and World Development Indicators Online) and publicly available sources (the Statistical Yearbook of the United Nations, World Bank Statistics, Eurostat Review, and Hofstede, 2001).

To assess innovative activity within the industry, we use patent statistics from the Delphion database (www.delphion.com). Delphion is a subscription-based database that contains detailed historical records on patents granted in the United States and other countries. Details on the measures used are provided in the next section.

3.2. Measures

This section describes the measure/operationalization of the DV, measures of the focal pricing constructs, measures of the moderator constructs, and measures of the remaining control variables.

3.2.1. Measure/operationalization of the DV-takeoff

Prior research uses threshold-based rules to guide the identification of takeoff adapted to suit the specific context (e.g., Golder & Tellis, 1997; Tellis & Stremersch, 2004; Tellis et al., 2003; Van Everdingen et al., 2009). In particular, Tellis et al. (2003) propose a measure of takeoff that is suitable for an international sample of countries. The authors define threshold as a standard plot of growth in sales for various levels of market penetration so that a more standard comparison of several countries can be obtained. Takeoff constitutes the first year in which an individual category’s growth rate relative to the base sales crosses this threshold. However, this rule allows takeoff to be identified using annual data. We modify this rule to suit our more temporally disaggregate data in the following manner.

We use the principle of bimonthly compounding to derive the growth rate of sales to suit bimonthly data regarding the annual growth rates proposed in Tellis et al. (2003). Hence, we use the following formula for compounded bimonthly growth:

$$r_2 = \left[\left(1 + \frac{r_1}{n_1}\right)^{\frac{n_2}{n_1}} - 1\right] n_2,$$

where

$$r_2$$ growth in sales at the bimonthly level

$$r_1$$ growth in sales at the annual level

$$n_2$$ 6 (6 bimonthly periods = 1 year)

$$n_1$$ 1 (to account for 1 year).

Fig. 4 displays the graph for the threshold for takeoff at the annual level, along with the corresponding bimonthly growth. Takeoff constitutes the first year in which the growth in sales for the new product exceeds the proposed growth threshold. Note that we consider takeoff to occur after a .5% threshold has been passed to avoid assessing takeoff at very low levels of market penetration.

We calculate the market penetration for all of the products in our database based on the following formula, where t refers to the bimonthly period:

$$Penetration_t = Penetration_{t-1} + \frac{sales_t}{households_t} \times 100.$$  

3.2.2. Measures of the focal pricing constructs

We calculate relative price as follows:

$$Relative\ price_t = price_t / price_{t-1}.$$  

As prices vary among products, this measure allows for standardization across products. Furthermore, relative price incorporates the use of the initial price of a new product as a reference point (Golder & Tellis, 1997; Rajendran & Tellis, 1994). Firms may intentionally decrease or raise prices based on the expected sales path (Agarwal & Bayus, 2002; Golder & Tellis, 1997; Kalish, 1983). We use a 1-period lagged measure of relative price to account for the potential endogeneity of price. The prices of all products are in US dollars.

We operationalize inter-temporal price volatility as follows. Consumers may form price expectations of the current price based on a price trend, as well as the most recently observed price. We measure inter-temporal price volatility based on the deviation of the actual and the expected price. Similar to Winer (1986), we use an extrapolative expectation model to determine the course of predicted prices in a market defined by country and category. Thus,

$$P_t = \delta_0 + \delta_1 P_{t-1} + \delta_2 \text{Time} + e_t,$$

where the subscript t refers to each bimonthly period. This equation is estimated separately for each group (product–country classification).

Eq. (4) assumes that a consumer’s prediction of the current price of a product is based on the most recent observed price (1 bimonthly period prior) and a time trend. Time is a control variable, which tracks the time period for the estimation of any trends in prices.

For each category–country combination, we estimate Eq. (4) for each of a series of moving windows of 4 bimonthly periods. For each period, starting from the fifth bimonthly period, we compute the Root Mean Squared Error (RMSE) of estimating Eq. (4), which provides an estimate of the volatility of prices within the moving window from the consumer’s perspective. We use a 1-period lagged measure in the regression.

3.2.3. Measures of moderator variables

Country-level macro-economic variables tend to be highly correlated (Dekimpe, Parker, & Sarvary, 2000). Hence, we operationalize wealth
by simply using national GDP per capita (in US dollars). We obtain the data for Real GDP per capita (Laspeyres) (Heston, Summers, & Aten, 2002). We use a 1 period lag for this measure.

We use a measure of uncertainty avoidance proposed by Hofstede (2001). This measure is available for all countries at http://www.geert-hofstede.com/.

We use a 1 period lagged measure of cumulative sales for a product in a country to capture intra-country contagion effects.

3.2.4. Measures of other control variables

For foreign takeoffs, we calculate the sum of prior takeoffs in the other countries (in our sample) within each of 4 regions (North America, Mid-Western Europe, Mediterranean Europe, and Nordic Europe). This method is consistent with prior research that suggests the existence of a regional influence on the hazard of takeoff (Tellis et al., 2003; Van Everdingen et al., 2009). We use a 1 period lagged measure in the regression.

We operationalize seasonality in the following manner. Although the use of patents as a measure of technological activity in an industry is somewhat problematic, as different firms may use different patenting strategies or exhibit varying levels of interest in obtaining patents, prior research has revealed a high correlation between level of patenting activity and the level of technological improvements in an industry (Bayus et al., 2007).

Fig. 4. Comparison of annual Tellis et al. (2003) rule and adapted bimonthly rule.

Seasonalityindex$=\sum_{j=1}^{n}\text{PMA}_{ij}/n$.

Fourth, we compute the average of the percent moving average for all products and years of observation (for a total of $j$ combinations) to obtain the seasonality index for each country and bimonthly period:

We use the seasonality index as an independent variable in the regression analysis. A positive coefficient implies that the hazard of takeoff is higher during peak seasons than during off-peak seasons.

3.3. Model

Because takeoff is a time-dependent event, we use a discrete time hazard model to test the hypotheses. The discrete time hazard approach allows for great flexibility in specifying the time function and in the incorporation of time-varying explanatory variables (Allison, 1995). The event is the takeoff of product $i$ in country $j$. The hazard of takeoff is the probability that product $i$ in country $j$ experiences takeoff during bimonthly period $t$, given the existence of no prior occurrences. If $T$ is an integer random variable providing the time of the event occurrence, conditional probability $P_{gi}$ (the probability that the event occurs at time $t$ given that it has not already occurred) is computed as follows:

$P_{gi} = \Pr \{ T = t \mid T \geq t, x_{gi} \}$.

where $x_{gi}$ is a vector of explanatory variables observed for product $i$ in country $j$ for period $t$. The model for the dependence of $P_{gi}$ on $x_{gi}$ can be specified as follows:

$P_{gi} = 1/(1 + e^{-\alpha - \beta x_{gi}})$.

This model becomes the following:

$\log \left( \frac{P_{gi}}{1 - P_{gi}} \right) = \alpha + \beta x_{gi}$.
For each product–country combination, there is a set of distinct observations, one for each unit of time until either the event (takeoff) occurs or the series is censored. For each of these observations, the dependent variable takes on a value of 1 if takeoff occurs during that time unit and a value of 0 otherwise. This method allows for the easy incorporation of time-varying covariates, corresponding to each period. Singer and Willet (1993) and Allison (1982) prove the equivalence between the likelihood of the discrete time hazard model and a sequence of N independent Bernoulli trials. Hence, maximum likelihood estimates can be derived using logistic regression in the discrete time framework. Thus, we pool these observations and estimate the following logistic regression model:

\[
\log\left(\frac{P_{ijt}}{1 - P_{ijt}}\right) = \alpha + \beta_1 \text{Relative Price}_{ij} + \beta_2 \text{Price Volatility}_{ij} + \beta_3 \text{National Wealth}_{jt} + \beta_4 \text{Uncertainty Avoidance}_{jt} + \beta_5 \text{Intra-country Contagion}_{ijt} + \beta_6 \text{Foreign Takeoff}_{ijt} + \beta_7 \text{Innovative Activity}_{jt} + \beta_8 \text{Seasonality}_{jt}.
\]

The proposed model (Eq. (12)) includes the main effects of both time-varying and time-invariant covariates. Time-varying covariates include relative price, price volatility, intra-country contagion effects, foreign takeoffs, innovative activity, seasonality, and national wealth (which is constant throughout a particular year but varies across years). The cultural dimension of uncertainty avoidance is time-invariant. Note that we do not include an effect for time because we capture the effect of contagion effects, measured as lagged cumulative sales.

We use the STATA logit procedure, allowing for the clustered robust standard error option, which specifies that the standard errors allow for intra-group (here, a product–country combination) correlation, relaxing the usual requirement that the observations be independent. That is, the observations are independent across groups (clusters) but not necessarily within groups.

4. Results

This section covers the identification of takeoff, other relevant descriptive statistics results obtained using our hazard model, and the robustness of the results.

4.1. Identification of takeoff

This section provides comparisons of the statistics on takeoff with regard to bimonthly and annual levels.

Table 2

<table>
<thead>
<tr>
<th>Description</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takeoff is not identified at either bimonthly level or annual level (product has not taken off)</td>
<td>12</td>
</tr>
<tr>
<td>Takeoff is identified at bimonthly level in the same year as rule identifies takeoff at annual level</td>
<td>73</td>
</tr>
<tr>
<td>Rule identifies takeoff at the bimonthly level in an earlier year than at annual level</td>
<td>15</td>
</tr>
</tbody>
</table>

We identify takeoff at the bimonthly level using the adapted Tellis et al. (2003) rule for 45 of the 51 product–country combinations. A visual inspection of penetration plots support the identification of takeoff for these product–country combinations. The average time until takeoff is 19.2 months, or roughly 2 years. This estimate is comparable to the 2 year interval for consumer electronics identified by Tellis et al. (2003) and the 3 year interval identified by Van Everdingen et al. (2009).

To assess the usefulness of measuring takeoff using bimonthly data, we apply the Tellis et al. (2003) rule using annual data and compare the results (Table 2). We obtain the annual sales data by aggregating across the 6 bimonthly periods. For instance, annual data for 1999 include the bimonthly periods of Feb–Mar, Apr–May, Jun–Jul, Aug–Sept, Oct–Nov, and Dec (99)–Jan (00). In some cases, bimonthly data end in Aug–Sept of the corresponding year. In all cases in which there are fewer bimonthly periods than would constitute one year, we estimate annual sales using proportional annualization.

We identify takeoff at the annual level for 42 out of 51 product–country combinations. In 12% of the cases, we do not identify takeoff at either the bimonthly or the annual levels, suggesting that the product has not yet taken off. In 73% of the cases, we identify takeoff using both bimonthly and annual data for the same year (this includes the cases in which the bimonthly data enable the identification of takeoff before the Dec–Jan period of that year). This result indicates that our adaptations of the Tellis et al. (2003) rule to the more disaggregate data level have face validity.

In 45% of the cases, we identify takeoff using the bimonthly rule during a month prior to the Dec–Jan bimonthly period in the same year. For example, in May 2003, takeoff is identified using the adapted bimonthly rule in June–July 2003 and, using annual data and the Tellis et al. (2003) rule, takeoff is identified in May 2003.

Furthermore, in 15% of the cases, we identify takeoff for an earlier year. The simple reason for this result is that there are 6 bimonthly data periods for every annual data period. This detail implies that with annual data, a greater number of observations is required for an accurate assessment and, although this type of data is more convenient to collect, we may not be able to pinpoint takeoff when it occurs over several months rather than several years.

4.2. Other key descriptive statistics for products and countries

On average and across countries, the least expensive products in our data are the MP3 player (audio) flash and audio HD. The most expensive products are the plasma TV and the LCD TV. Furthermore, audio flash and audio HD pricing are relatively less volatile than that for the plasma and the LCD TV (Table 3a).

On average, across products and countries, we find that prices decrease substantially during the early life cycles of new products.

\footnote{That is, for each product–country combination, we first determine the proportion of total sales throughout the entire period of observation, which is contributed to by each bimonthly period (if for the year 2005, data for only 2 bimonthly periods are present, we estimate the annual sales in 2005 by dividing the total sales over the 2 bimonthly periods by the proportion of sales represented by the two bimonthly periods).}
On average, prices are 5% lower than launch prices 1 year after commercialization. After 2 years, prices are 27% lower than the launch prices. After 3–4 years, prices are approximately 50% of the corresponding launch prices. On average, across products and countries, prices at takeoff are 52% of initial prices.

The countries represent a spectrum of uncertainty avoidance scores that range from low (Sweden, the United Kingdom and the United States) to high (Spain, France and Italy). Although our sample consists of only developed countries, the United States may be characterized as among the wealthier countries in the dataset, whereas Spain may be characterized as the least wealthy country in the dataset (see Table 3b).

We next examine the factors influencing the takeoff of new products using data at the disaggregate level and in an international context.

### 4.3. Estimates of the hazard model: main effects

The correlations among the independent variables are in Table 4. The estimates of the discrete time hazard model (Eq. (12)) are in Table 5. We cluster according to product–country combination and employ robust standard errors.

Model 1 is the base model without the interaction effects. We find that relative price has a negative and significant effect. Hence, lower relative prices increase the hazard of takeoff. These results support H1. Hence, across products and countries, we find that relative price affects the takeoff of new products.

We also find that price volatility has a negative and significant coefficient. Hence, we obtain support for H2, which holds that higher price volatility decreases the hazard of takeoff.

With regard to the non-price dimensions, we determine that higher levels of intra-country contagion levels and innovative activity and higher numbers of foreign takeoffs increase the hazard of takeoff. Higher uncertainty avoidance decreases the hazard of takeoff and, unexpectedly, higher wealth (GDP per capita) does as well.

The coefficient of seasonality is positive and significantly different from 0. Hence, the hazard of takeoff is higher during peak seasons.

To assess the models’ fit, we use the pseudo R-square. Our results for Model 1, in Table 5, reveal a pseudo R-square of .34, which is consistent with prior research (Golder & Tellis, 1997, who report an R-square like measure of .31; Tellis et al., 2003 report .18 for the full model).

When independent variables are measured at different scales or in different units, beta coefficients must be standardized so that the relative effects of the independent variables can be compared. We derive the standardized estimates using the STB option in SAS Proc Logistic. The standardized coefficients for the main-effects only model are in Table 6. The interpretation of these standardized coefficients is as follows: A one-standard deviation increase in the independent variable (X) produces a b* standard deviation change in the logit of (Y). For example, an increase of one standard deviation in innovative activity is associated with an increase of .23 standard deviations in the logit of takeoff. A one-standard deviation increase in relative price leads to a decrease of .43 standard deviations in the logit of takeoff. Thus, in terms of the direct effects, we find that price volatility

### Table 3a

Descriptive statistics on pricing dimensions and takeoff patterns by product.

<table>
<thead>
<tr>
<th>Product</th>
<th>#a</th>
<th># Takeoffs</th>
<th>Mean price at commercialization</th>
<th>Mean relative price at takeoffb</th>
<th>Mean price volatility at one period before takeoff (rank 1 = low, 7 = high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVD recorder</td>
<td>8</td>
<td>8</td>
<td>2206.66</td>
<td>.21</td>
<td>3</td>
</tr>
<tr>
<td>Digital video recorder</td>
<td>6</td>
<td>2</td>
<td>1427.90</td>
<td>.44</td>
<td>5</td>
</tr>
<tr>
<td>Home theater</td>
<td>8</td>
<td>8</td>
<td>906.44</td>
<td>.76</td>
<td>4</td>
</tr>
<tr>
<td>Plasma</td>
<td>7</td>
<td>6</td>
<td>14876.33</td>
<td>.22</td>
<td>7</td>
</tr>
<tr>
<td>LCD</td>
<td>8</td>
<td>8</td>
<td>2807.10</td>
<td>.54</td>
<td>6</td>
</tr>
<tr>
<td>Audio flash</td>
<td>7</td>
<td>7</td>
<td>219.71</td>
<td>.65</td>
<td>1</td>
</tr>
<tr>
<td>Audio HD</td>
<td>7</td>
<td>6</td>
<td>412.40</td>
<td>.75</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Prices in US $.

a Number of countries for which data on this category is available.

b Calculated for cases where takeoff has occurred.
has the largest effect, followed by, in descending order, wealth, intra-country interactions, relative price, uncertainty avoidance, seasonality, innovative activity, and foreign takeoffs.

4.4. Estimates of the hazard model: interaction effects

We re-estimate Eq. (12) after including the interaction terms between the pricing dimensions and wealth, culture, and contagion. We sequentially include blocks of the interaction terms to facilitate interpretation.

Model 2 adds the interaction term between the pricing dimensions and national wealth. The results of this analysis are in Table 5. We do not find a significant effect for the interaction between relative price and national wealth. Hence, we do not obtain support for H3. However, we find that the interaction between price volatility and national wealth has a positive and significant effect. This result implies that in countries with high national wealth, the influence of price volatility on takeoff becomes less important. This finding supports H4.

Model 3 adds the interaction term between the pricing dimensions and uncertainty avoidance. We do not find a significant interaction effect between relative price and uncertainty avoidance. Hence, the results do not support H5. However, we find that the interaction between price volatility and uncertainty avoidance has a negative and significant effect. This result implies that the negative influence of price volatility on takeoff is enhanced in countries with high uncertainty avoidance, in support of H6.

Model 4 adds the interaction term between the pricing dimensions and intra-country contagion effects. We do not find a significant interaction between relative price and intra-country contagion effects. However, we find that the interaction between price volatility and intra-country contagion effects has a positive and significant effect. The results from Model 4 therefore do not support H7, but do support H8. This result implies that for products associated with high intra-country contagion effects, price volatility’s effect on takeoff decreases.

4.5. Out-of-sample prediction

Can managers use our model to predict takeoff? We use a jackknife method to simulate the context of the manager of a target product in a target country. The jackknife method ascertains the out-of-sample predictive validity of the hazard model. We re-estimate Eq. (12) (plus the interaction terms) n times, excluding one market each time. Here, n is the number of markets (product-country combinations) in our sample. For each of these n runs, we use the estimated parameters of the model to predict the hazard of takeoff for the excluded target market. We compare predicted and actual takeoffs across these n iterations. We calculate the model’s accuracy in predicting takeoff in terms of various hit rates. Specificity is the power of the model to detect true negatives, whereas sensitivity is the power of the model to detect true positives (see Eqs. (13) and (14)).

\[
\text{Specificity} = \frac{\text{TrueNegatives}}{\text{ActualNegatives}} = \frac{\text{TrueNegatives}}{\text{TrueNegatives} + \text{FalsePositives}} \quad (13)
\]

\[
\text{Sensitivity} = \frac{\text{TruePositives}}{\text{ActualPositives}} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \quad (14)
\]

There is a trade-off between sensitivity and specificity. This distinction is sensitive to the relative size of the component groups and favors classification in the larger group (Hosmer & Lemeshow, 2004). In our case, there are many more occurrences of 0s than 1s and, thus, we choose an optimal cut-off point that maximizes both sensitivity and specificity. For the n iterations, model sensitivity is 61% and specificity is 91%, which is much higher than what could have been predicted by chance alone.

4.6. Robustness analysis

We next examine the robustness of the above analysis to the following: the inclusion of measures of industry level competition, additional pricing dimensions, an alternative hazard model specification, different control variables, a different price volatility specification, and the inclusion of different clustering specifications.

4.6.1. Impact of competitive activity

Competitive activity in the industry may affect product takeoff (Bayus et al., 2007). An increase in competitive activity denotes the entry of several new players into the market, who offer newer versions of the product, which may stimulate takeoff (Agarwal & Bayus, 2002). Research posts that tacit collusion to thwart new entrants is easier in markets characterized by high levels of concentration.

---

**Table 3b**

Descriptive statistics on pricing dimensions and takeoff patterns by country.

<table>
<thead>
<tr>
<th>Country</th>
<th># Takeoffs</th>
<th>Mean relative price at takeoff</th>
<th>Uncertainty avoidance scores</th>
<th>Mean GDP per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>6</td>
<td>0.52</td>
<td>46</td>
<td>34,544</td>
</tr>
<tr>
<td>UK</td>
<td>7</td>
<td>0.56</td>
<td>35</td>
<td>23,297</td>
</tr>
<tr>
<td>Sweden</td>
<td>4</td>
<td>0.41</td>
<td>29</td>
<td>24,733</td>
</tr>
<tr>
<td>Spain</td>
<td>7</td>
<td>0.50</td>
<td>86</td>
<td>19,061</td>
</tr>
<tr>
<td>Netherlands</td>
<td>7</td>
<td>0.65</td>
<td>53</td>
<td>25,525</td>
</tr>
<tr>
<td>Italy</td>
<td>7</td>
<td>0.42</td>
<td>75</td>
<td>22,561</td>
</tr>
<tr>
<td>Germany</td>
<td>7</td>
<td>0.49</td>
<td>65</td>
<td>23,488</td>
</tr>
<tr>
<td>France</td>
<td>6</td>
<td>0.52</td>
<td>86</td>
<td>23,034</td>
</tr>
</tbody>
</table>

Note: Prices in US $.

- a Number of categories for which data is available within each country.
- b Calculated for cases where takeoff has occurred.
- c Average GPPPC estimated across years for each country.

**Table 4**

Correlation matrix.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Takeoff</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Relative price (lag 1)</td>
<td>-0.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Price volatility (lag 1)</td>
<td>-0.08</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Innovative activity (lag 1)</td>
<td>0.14</td>
<td>0.27</td>
<td>-0.13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>National wealth (lag 1)</td>
<td>0.04</td>
<td>-0.11</td>
<td>-0.11</td>
<td>0.01</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Uncertainty avoidance</td>
<td>-0.02</td>
<td>0.1</td>
<td>-0.09</td>
<td>0.01</td>
<td>-0.52</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Seasonality</td>
<td>0.05</td>
<td>-0.06</td>
<td>0</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>Foreign takeoff (lag 1)</td>
<td>0.33</td>
<td>-0.17</td>
<td>-0.09</td>
<td>0.17</td>
<td>-0.07</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>9</td>
<td>Contagion effects (lag 1)</td>
<td>0.32</td>
<td>-0.21</td>
<td>-0.15</td>
<td>0.11</td>
<td>0.50</td>
<td>-0.11</td>
<td>-0.00</td>
</tr>
</tbody>
</table>
pany market shares from 2002 to 2006 in terms of percentages of
Global Marketing Information Database). We are able to obtain com-
market shares from the Euromonitor country reports (Euromonitor
data set and for each product market, we obtain data on company
lower the level of competition are. For each of the 8 countries in our
is to being a monopoly, the higher the market’s concentration and the
Hirshman Index is derived by calculating the sum of the squares of
competitive activity in an industry is an important control variable.
Hence, the extent of
purchase of new consumer packaged goods. Hence, the extent of
Notes: Observations: 615, robust z-statistics in parentheses, *** p
260
Table 6
Results from hazard model (drivers of takeoff).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative price (lag 1)</td>
<td>H1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price volatility (lag 1)</td>
<td>H2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National wealth (lag 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertainty avoidance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contagion (lag 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign takeoffs (lag 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovative activity (lag 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seasonality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price (lag 1) * National wealth (lag 1)</td>
<td>H3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price volatility (lag 1) * National wealth (lag 1)</td>
<td>H4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price (lag 1) * Uncertainty avoidance</td>
<td>H5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price volatility (lag 1) * Uncertainty avoidance</td>
<td>H6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative price (lag 1) * Contagion (lag 1)</td>
<td>H7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price volatility (lag 1) * Contagion (lag 1)</td>
<td>H8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>5.253**</td>
<td>8.703***</td>
<td>5.757**</td>
<td>4.596</td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.339</td>
<td>0.353</td>
<td>0.366</td>
<td>0.373</td>
</tr>
</tbody>
</table>

Notes: Observations: 615, robust z-statistics in parentheses, *** p < 0.01, ** p < 0.05, and * p < 0.1, 2-sided significance levels.

Gielens and Steenkamp (2007) report that increasing concentration negatively impacts both first year purchases and trends in the purchase of new consumer packaged goods. Hence, the extent of competitive activity in an industry is an important control variable.
To control for this effect, we develop a measure of the level of the concentration in the industry. We measure the Herfindahl–Hirschman Index of concentration in each product market. The Herfindahl–Hirschman Index is derived by calculating the sum of the squares of the market shares of every firm in the industry. The closer the market is to being a monopoly, the higher the market’s concentration and the lower the level of competition are. For each of the 8 countries in our data set and for each product market, we obtain data on company market shares from the Euromonitor country reports (Euromonitor Global Marketing Information Database). We are able to obtain company market shares from 2002 to 2006 in terms of percentages of retail volume in the following product categories: home audio and video, portable media players, TV systems & projectors, and video/DVD systems.

We next run a separate hazard analysis by including this variable as an independent variable in the regression, along with 1 period lag, excluding the years before 2002. We find that the coefficient for the Herfindahl index is not significantly different from 0, whereas the other results remain substantially robust.

4.6.2 Additional pricing dimensions
We test the robustness of the model to the inclusion of additional pricing dimensions. We examine whether takeoff may be impacted by either the magnitude of the price increase or the magnitude of the price decrease during the previous period for each product market. We include a variable assessing price decrease...
(increase), 1 period lagged, in the regression. We find that there is no significant increase (decrease) in the hazard of takeoff for this additional pricing dimension, whereas the other results remain robust.

We also consider the robustness of the analysis to the inclusion of absolute prices rather than relative prices. When absolute price is considered in Model 1, either the pricing dimensions are not significant or only the absolute price is significant (at the 10% level). There are issues with considering the absolute price. First, the reference price literature, as well as the underlying rationale based on economic theory, suggests that price relative to some threshold rather than absolute price, plays a role. Second, we obtain high correlations of 0.41 with the price volatility measure (as well as the measure we use in our robustness checks), whereas the correlation between our relative price measure and price volatility is only 0.05.

### 4.6.3. Alternative hazard model specification

We examine the robustness of our analysis to an alternate complementary log–log hazard specification. We find that our analysis is robust to all of the results for the hypotheses, with the exception of H4, on the interaction between price volatility and national wealth, for which we find no support.

### 4.6.4. Additional control variables

We test the robustness of our analysis to the inclusion of additional control variables tested in the literature on takeoff. Models 1 and 2 in Table A of the Appendix A show the impact of the addition of income inequality, as well as the three other dimensions of Hofstede’s dimensions of culture. Our key findings remain robust. Models 3 and 4 show the inclusion of time since commercialization. Our results remain substantially robust (the results do not support the interaction of price volatility and GDP per capita).

### 4.6.5. Alternate price volatility measure

We consider a different specification of the price volatility measure to capture the impact of an exponentially decreasing price, such as the following:

\[
P_t' = \delta_0 + \delta_1 \text{Time} + \delta_2 \text{Time}^2 + \epsilon_{gt},
\]

where subscript t refers to the time period. Eq. (15) assumes that a consumer’s prediction of the current price of a product is based on an exponential specification of Time, where Time tracks the time period, beginning at commercialization. This equation is estimated separately for each group (product–country classification). Our results remain robust, with the exception of the interaction effect between price volatility and uncertainty avoidance, which is no longer significantly different from 0 (see Appendix Table B).

### 4.6.6. Alternate clustering specifications

We allow for clustered error terms in small clusters (same product and same cluster). We determine the robustness to 2 other clustering specifications. One specification allows for the clustering of all observations addressing the same product across countries (Models 1 and 2 in Appendix Table C). The other specification allows for correlations among all observations across products belonging to the same country (Models 3 and 4 in Appendix Table C). The results are robust to these alternate specifications.

### 5. Discussion

New product takeoff is a critical event in the life cycle of new products, potentially signaling mass adoption and ultimate success. Managing the drivers that may affect takeoff is therefore of crucial importance for managers. Pricing is one of the most important marketing variables that practitioners use to manage new products (e.g., Gijsbrechts, 1993). However, various dimensions of price have not been considered in-depth with regard to the international takeoff of new products. This study focuses on the effect of two dimensions of price (relative price and price volatility) on the international takeoff of new products using a novel data set of bimonthly observations of 7 new consumer electronics products in 8 countries. This section discusses the key results, implications, and limitations of the study.

#### 5.1. Key results

The key results of the study are as follows:

- Both relative price and price volatility significantly impact the hazard of takeoff. Price volatility may be one of the strongest factors driving (or hindering) takeoff.
- However, although the influence of price decreases relative to introductory price is stable across contexts, the effect of price volatility is moderated by wealth, culture, and contagion. The effect of price volatility is enhanced in countries with high uncertainty avoidance and diminished in countries with greater wealth. Moreover, for products that have high intra-country contagion effects, price volatility has a reduced impact.
- Greater data granularity leads to the identification of inter-temporal effects such as price volatility, which is a critical determinant of takeoff.

#### 5.2. Implications

Our results have the following implications.

First, managers must carefully monitor prices to manage and accelerate takeoff. For takeoff, prices must eventually be substantially lower than introductory prices, strongly suggesting the need for a skimming strategy. Thus, managers should carefully consider the level of the introductory price and plan for regular price decreases.

Second, higher price volatility significantly lowers the hazard of takeoff. In fact, our analysis reveals that price volatility may be one of the important variables hindering takeoff, and an inability to obtain data at a disaggregate level may be seriously impacting our understanding of volatility’s effects on takeoff. While executing a price skimming strategy, managers must plan to use this strategy judiciously, implementing steady rather than erratic price drops. Utilizing a steady price decrease strategy is particularly relevant in less wealthy or less venturesome countries, which are more sensitive to price volatility. In addition, although cost and competitive pressures may cause price volatility, managers must minimize such volatility to avoid the delay or abortion of a new product takeoff.

Third, earlier studies highlight the importance of managing social contagion during the growth and maturity stages (e.g., Chandrasekaran & Tellis, 2011; Goldenberg, Libai, & Muller, 2002; Prees et al., 2010; Van den Bulte & Stremersch, 2004). Our results highlight the importance of social contagion to the stimulation of takeoff at an earlier stage of the product life cycle. Furthermore, although some reviews speculate regarding a trade-off in terms of managerial time in which there is a focus on pricing rather than contagion (e.g., Peres et al., 2010), we find evidence of a positive interplay between pricing and contagion: increased social contagion may diminish price volatility’s influence on takeoff.

Fourth, managers must know as early as possible whether takeoff occurs, a necessity that has gained importance in recent times, as takeoff appears to occur quickly, within months. This study provides a method for assessing takeoff at the bimonthly level through the adaptation of the Tellis et al. (2003) rule. Thus, in the context of bimonthly data, a manager can determine at the end of each 2-month period whether takeoff has occurred. Assessing takeoff using
temporally disaggregate data can enable the pinpointing of takeoff within the year. This method will provide firms with control and flexibility for managing new products.

5.3. Limitations

This study has several limitations that suggest opportunities for future research. First, our research focuses on product categories, not individual brands. Thus, we consider average price for a category rather than the prices of individual brands. This practice is common in research on takeoff and diffusion because of the difficulty of procuring brand-level data across countries. Second, although this study is based on a set of multiple products and countries, using a more comprehensive set of products and countries will produce more generalizable findings. Third, we measure firm entry across countries according to industry concentration, which reflects the number of sellers in a market. Data on numbers of entrants would be very useful. Fourth, although our data are bimonthly, they are aggregated at the country level. Assessing takeoff using cross-sectional disaggregate data would be advantageous. Fifth, we do not consider the effects of the prices of competing products. All of these issues remain promising areas for future research.

Acknowledgments

The authors thank participants of the Marketing Science Conference, Singapore for their comments. The authors also thank Tammo Bijmolt and Stefan Stremersch for their insightful comments and suggestions on earlier drafts of this paper. This research was supported by a grant from Don Murray to the USC Center for Global Innovation.

Appendix A

Table A

<table>
<thead>
<tr>
<th></th>
<th>(1)(a)</th>
<th>(2)(a)</th>
<th>(3)(b)</th>
<th>(4)(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative price (lag 1)</td>
<td>-3.468***</td>
<td>4.875</td>
<td>-2.474**</td>
<td>7.393</td>
</tr>
<tr>
<td>Price volatility (lag 1)</td>
<td>-0.00766***</td>
<td>0.197**</td>
<td>-0.0112***</td>
<td>-0.0829</td>
</tr>
<tr>
<td>Income inequality (lag 1)</td>
<td>0.0853</td>
<td>0.0917</td>
<td>0.112</td>
<td>0.133</td>
</tr>
<tr>
<td>National wealth (lag 1)</td>
<td>-0.00126***</td>
<td>-0.00139***</td>
<td>-0.00140***</td>
<td>-0.00143***</td>
</tr>
<tr>
<td>Uncertainty avoidance</td>
<td>0.0513</td>
<td>0.106</td>
<td>0.0663</td>
<td>0.133*</td>
</tr>
<tr>
<td>Power distance</td>
<td>-0.0496</td>
<td>-0.0489</td>
<td>-0.0782</td>
<td>-0.0897</td>
</tr>
<tr>
<td>Masculinity</td>
<td>-0.134***</td>
<td>-0.155***</td>
<td>-0.148***</td>
<td>-0.171***</td>
</tr>
<tr>
<td>Individualism</td>
<td>0.245***</td>
<td>0.274***</td>
<td>0.267***</td>
<td>0.297***</td>
</tr>
<tr>
<td>Contagion (lag 1)</td>
<td>3.68e-05***</td>
<td>3.94e-05***</td>
<td>3.97e-05***</td>
<td>4.04e-05***</td>
</tr>
<tr>
<td>Foreign takeoff (lag 1)</td>
<td>1.060**</td>
<td>1.283**</td>
<td>1.047**</td>
<td>1.576**</td>
</tr>
<tr>
<td>Innovative activity (lag 1)</td>
<td>0.0110**</td>
<td>0.0115**</td>
<td>0.0122**</td>
<td>0.0125**</td>
</tr>
<tr>
<td>Seasonality</td>
<td>1.682***</td>
<td>1.900***</td>
<td>1.608***</td>
<td>1.774***</td>
</tr>
<tr>
<td>Relative price (lag 1) * National wealth (lag 1)</td>
<td>-0.000219</td>
<td>-0.000229</td>
<td>-0.000298</td>
<td>-0.000298</td>
</tr>
<tr>
<td>Price volatility (lag 1) * National wealth (lag 1)</td>
<td>5.20e-05**</td>
<td>5.08e-05**</td>
<td>5.91e-05**</td>
<td>6.10e-05**</td>
</tr>
<tr>
<td>Relative price (lag 1) * Uncertainty avoidance</td>
<td>-0.0495</td>
<td>-0.0474</td>
<td>-0.0474</td>
<td>-0.0474</td>
</tr>
<tr>
<td>Price volatility (lag 1) * Uncertainty avoidance</td>
<td>-0.000699***</td>
<td>-0.000769***</td>
<td>-0.000769***</td>
<td>-0.000769***</td>
</tr>
<tr>
<td>Relative price (lag 1) * Contagion (lag 1)</td>
<td>-3.97e-07</td>
<td>-3.97e-07</td>
<td>-3.97e-07</td>
<td>-3.97e-07</td>
</tr>
<tr>
<td>Price volatility (lag 1) * Contagion (lag 1)</td>
<td>1.74e-07***</td>
<td>1.74e-07***</td>
<td>1.67e-07***</td>
<td>1.67e-07***</td>
</tr>
<tr>
<td>Time since commercialization</td>
<td>0.0895***</td>
<td>0.0876***</td>
<td>0.0876***</td>
<td>0.0876***</td>
</tr>
<tr>
<td>Constant</td>
<td>9.311***</td>
<td>7.286</td>
<td>9.106**</td>
<td>4.318</td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.474</td>
<td>0.522</td>
<td>0.500</td>
<td>0.542</td>
</tr>
</tbody>
</table>

Notes: Observations: 615, Robust z-statistics in parentheses, *** p < 0.01, ** p < 0.05, and * p < 0.1, 2-sided significance levels.

\(a\) Model shows inclusion of income inequality and cultural variables.

\(b\) Model shows inclusion of additional variable-time since commercialization.
Table B
Robustness checks — inclusion of a different specification of price volatility.

<table>
<thead>
<tr>
<th></th>
<th>(1)*</th>
<th>(2)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative price (lag 1)</td>
<td>−2.286***</td>
<td>2.258</td>
</tr>
<tr>
<td>Price volatility (lag 1)</td>
<td>−0.00555**</td>
<td>−0.0672**</td>
</tr>
<tr>
<td>National wealth (lag 1)</td>
<td>−0.000300***</td>
<td>−0.000308</td>
</tr>
<tr>
<td>Uncertainty avoidance</td>
<td>−0.0257**</td>
<td>0.000592</td>
</tr>
<tr>
<td>Contagion (lag 1)</td>
<td>1.42e − 05***</td>
<td>9.14e − 06</td>
</tr>
<tr>
<td>Foreign takeoffs (lag 1)</td>
<td>0.715**</td>
<td>1.269***</td>
</tr>
<tr>
<td>Innovative activity (lag 1)</td>
<td>0.0103**</td>
<td>0.0108**</td>
</tr>
<tr>
<td>Seasonality</td>
<td>1.305***</td>
<td>3.21**</td>
</tr>
</tbody>
</table>

Relative price (lag 1) × National wealth (lag 1)

<table>
<thead>
<tr>
<th></th>
<th>(1)*</th>
<th>(2)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price volatility (lag 1)</td>
<td>−2.43e − 06***</td>
<td>0.000105</td>
</tr>
<tr>
<td>Relative price (lag 1) × Uncertainty avoidance</td>
<td>−0.0407</td>
<td>0.0151</td>
</tr>
<tr>
<td>Price volatility (lag 1) × Uncertainty avoidance</td>
<td>−0.000308</td>
<td>0.0101</td>
</tr>
<tr>
<td>Relative price (lag 1) × Contagion (lag 1)</td>
<td>5.23e − 06</td>
<td>1.76e − 07**</td>
</tr>
<tr>
<td>Price volatility (lag 1) × Contagion (lag 1)</td>
<td>4.67e**</td>
<td>3.639</td>
</tr>
</tbody>
</table>

Psuedo R-square | 0.333 | 0.364 |

Notes: Observations: 662, Robust z-statistics in parentheses, *** p < 0.01, ** p < 0.05, and * p < 0.1, 2-sided significance levels.
a Model shows inclusion of a different price volatility measure.

Table C
Robustness checks — inclusion of alternate clustering specifications.

<table>
<thead>
<tr>
<th></th>
<th>(1)*</th>
<th>(2)*</th>
<th>(3)*</th>
<th>(4)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative price (lag 1)</td>
<td>−2.733***</td>
<td>−1.320</td>
<td>−2.733***</td>
<td>−1.320</td>
</tr>
<tr>
<td>Price volatility (lag 1)</td>
<td>−0.00602***</td>
<td>−0.0314**</td>
<td>−0.00602**</td>
<td>−0.0314**</td>
</tr>
<tr>
<td>National wealth (lag 1)</td>
<td>−0.000308***</td>
<td>−0.000384**</td>
<td>−0.000308*</td>
<td>−0.000384*</td>
</tr>
<tr>
<td>Uncertainty avoidance</td>
<td>−0.0278***</td>
<td>0.0151</td>
<td>−0.0278</td>
<td>0.0151</td>
</tr>
<tr>
<td>Contagion (lag 1)</td>
<td>1.36e − 05***</td>
<td>1.36e − 05***</td>
<td>1.36e − 05***</td>
<td>1.36e − 05***</td>
</tr>
<tr>
<td>Prior takeoff (lag 1)</td>
<td>0.708***</td>
<td>0.633**</td>
<td>0.708</td>
<td>0.633</td>
</tr>
<tr>
<td>Innovative activity (lag 1)</td>
<td>0.01096***</td>
<td>0.01355***</td>
<td>0.01096***</td>
<td>0.01355***</td>
</tr>
<tr>
<td>Seasonality</td>
<td>1.274***</td>
<td>1.274***</td>
<td>1.274***</td>
<td>1.274***</td>
</tr>
</tbody>
</table>

Relative price (lag 1) × National wealth (lag 1)

<table>
<thead>
<tr>
<th></th>
<th>(1)*</th>
<th>(2)*</th>
<th>(3)*</th>
<th>(4)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price volatility (lag 1) × National wealth (lag 1)</td>
<td>5.55e − 05</td>
<td>5.55e − 05</td>
<td>5.55e − 05</td>
<td>5.55e − 05</td>
</tr>
<tr>
<td>Relative price (lag 1) × Uncertainty avoidance</td>
<td>−0.0426</td>
<td>−0.0426</td>
<td>−0.0426</td>
<td>−0.0426</td>
</tr>
<tr>
<td>Price volatility (lag 1) × Uncertainty avoidance</td>
<td>−0.000470***</td>
<td>−0.000470**</td>
<td>−0.000470**</td>
<td>−0.000470**</td>
</tr>
<tr>
<td>Relative price (lag 1) × Contagion (lag 1)</td>
<td>2.44e − 07</td>
<td>2.44e − 07</td>
<td>2.44e − 07</td>
<td>2.44e − 07</td>
</tr>
<tr>
<td>Price volatility (lag 1) × Contagion (lag 1)</td>
<td>7.23e − 08***</td>
<td>7.23e − 08***</td>
<td>7.23e − 08***</td>
<td>7.23e − 08***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)*</th>
<th>(2)*</th>
<th>(3)*</th>
<th>(4)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.253***</td>
<td>5.253</td>
<td>5.253</td>
<td>5.253</td>
</tr>
<tr>
<td>Psuedo R-square</td>
<td>0.339</td>
<td>0.373</td>
<td>0.339</td>
<td>0.373</td>
</tr>
</tbody>
</table>

Notes: Observations: 615, Robust z-statistics in parentheses, *** p < 0.01, ** p < 0.05, and * p < 0.1, 2-sided significance levels.
a Model includes clustering of all observations dealing with same product across countries.
b Model includes clustering of all observations across products belonging to same country.
References


Is power powerful? Power, confidence, and goal pursuit

Dongwon Min a,1, Ji-Hern Kim b,*

a College of Business and Economics, Dankook University, 125, Jukjeon-dong, Su-ji, Yong-in, Gyeonggi-do, 448-701, South Korea
b School of Business, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul, 143-747, South Korea

ARTICLE INFO

Article history:
First received in 17, August 2011
and was under review for 8½ months
Available online 2 April 2013
Area Editor: Marnik G. Dekimpe

Keywords:
Power
Confidence
Goal pursuit
Attention
Recall
Persuasion

1. Introduction

How can one develop goals or beliefs based on the infinite amount of information available in the social world? How can one process and understand such information? In recent years, researchers have focused on power as an essential force controlling one’s mental resources (Galinsky, Gruenfeld, & Magee, 2003; Kipnis, 1972). Power has long been demonstrated to be a fundamental variable in social structures. In the past three decades, social psychologists have devoted much theoretical and empirical attention to how the experience of holding power impacts various and significant psychological phenomena in human behavior. Research has shown that power leads people to rely on stereotypic information to maintain their status quo (Fiske & Deppret, 1996) and to use less costly information processing strategies, such as heuristic processing (Fiske, 1993) and automatic cognition (Galinsky et al., 2003). In studies relating more directly to the effects of social power on social motivation, findings have supported the idea that power reinforces flexible and selective information processing. Specifically, relative to the powerless, the powerful pay less attention to others and to counter-stereotypic (individualizing) information (Fiske, 1993; Goodwin, Gubin, Fiske, & Yzerbyt, 2000), pay more attention to information that confirms their expectations (Ebenbach & Keltner, 1998), and more extensively process information relevant to accessible constructs, such as goals or needs (Guinote, 2007a). Although researchers have long recognized the value of psychological power as an essential stimulating force in cognitive processing within various social contexts, the relationship of psychological power to marketing has received little attention.

The present study investigates how power influences cognitive processing in an environment where power is irrelevant. Specifically, we focus on the role of power perception in processing a persuasive message. Based on previous studies on power, confidence, and goal pursuit, we predict that the powerful will pay less attention to such a message and thus should exhibit poorer recall performance than those with low power; confidence that is induced by power mediates the effect of power. In contrast, when a specific goal is established, those with high power are better at processing goal-relevant messages than those with low power; confidence does not play a mediating role in this condition (Experiment 3).

© 2013 Elsevier B.V. All rights reserved.

Abstract

We investigate the influence of power on the cognitive processing of persuasive messages by examining how people with high power pay attention to recall, and are persuaded by messages relative to those with low power. We employ multiple power manipulations by placing participants in a hierarchical structure (Experiment 1) or priming them by asking them to recall an event in which they either had power over someone or someone had power over them (Experiments 2 and 3). The results reveal that, in a neutral setting, those with high power perform worse on attention and recall and are less persuaded by a message than those with low power; confidence that is induced by power mediates the effect of power. In contrast, when a specific goal is established, those with high power are better at processing goal-relevant messages than those with low power; confidence does not play a mediating role in this condition (Experiment 3).
2. Theoretical background

2.1. What is power?

Previous studies have defined psychological power in many ways. Some defined it as a basic force underlying social relationships (Fiske, 1993; Kemper, 1981) and behaviors (Emerson, 1962; Turner, 2005). A growing number of studies have examined the effects of power as a cognitive structural variable and found that the powerful are less likely than the powerless to pay attention to others (Fiske, 1993; Keltner & Robinson, 1997) and to information that disconfirms their expectations (Ebenbach & Keltner, 1998). These studies, which are more directly related to the relationship between power and cognitive efforts, revealed that power reduced deliberation in processing and indicated that the powerful tended to use simpler cognitive strategies (Gruenfeld, 1995) and evaluated target individuals as displaying less cognitive complexity (Woike, 1994). Related to this issue, Fiske (1993) documented that, because the outcomes of the powerful depend less on others than those of the powerless, the powerful should have less motivation to be accurate in their estimates of others and be more likely to use cognitively inexpensive strategies. Recent studies have confirmed the proposed link between power and less effortful processing strategies (e.g., Chen, Ybarra, & Kiefer, 2004): the powerful tended to process the world in a less effortful, less deliberate, and more heuristic manner than the powerless. Specifically, such findings indicated that power increased an individual’s propensity to use less systematic cognition (Gruenfeld, 1995) but decreased the perceived need to attend to and to process subsequent information (Britol et al., 2007).

Through various investigations, Galinsky et al. (2003: 453) argued that “power can be conceived not only as an aspect of social structure, but also as a cognitive structure that can be activated by an appropriate environmental stimulus”. According to their findings, a new situation entailing the possession of power, or an individual’s simple recollection of an experience with power, induced the behavioral tendencies associated with power (e.g., Galinsky, Magee, Gruenfeld, Whitson, & Liljenquist, 2008; Galinsky et al., 2003). This can be explained by the idea of Bargh, Raymond, Pryor, and Strack (1995) and Chen, Lee-Chai, and Bargh (2001) that the activation of power as a construct also activates concepts associated with power, which in turn increases the behavioral tendencies relating to power. Based on these findings, we provide an integrative account of the effects of power as a cognitive structure. Specifically, we focus on social power, which is derived from one’s relationships with others (Overbeck & Park, 2001), and its role in attention, memory, and persuasion.

2.2. Power and confidence

To understand the influence of power on cognitive processing, we need to determine what drives the effects of power. Previous studies on power have posited a close link between power and confidence (Anderson & Galinsky, 2006; Magee, Milliken, & Lurie, 2010), which can be defined as a “subjective sense of conviction about one’s beliefs and opinions” (Briol et al., 2007: 1041). These studies offered evidence that the powerful are more likely to act confidently than the powerless and that feeling powerful leads individuals to rely on their sense of confidence. Specifically, Briol et al. (2007) demonstrated that power influenced persuasion differently depending on when the power was induced. According to their findings, when power was induced prior to a message, it led to the validation of one’s existing views, whereas when power was induced after a message, it increased confidence in one’s thoughts. Thus, the increased confidence validated one’s recently generated thought. As a result, individuals whose high power was induced after a message were less likely to differentiate between strong and weak arguments than those with low power. The authors concluded that power induced after a message increased confidence in one’s thoughts, which validated the thought in one’s mind, and thus reduced the amount of information one can process. Consistent with this idea, a number of studies examined the effects of confidence on information processing and judgment (e.g., Tiedens & Linton, 2001; Weary & Jacobson, 2007), demonstrating that confidence had a considerable influence on the extent to which individuals engage in more effortful systematic processing versus less effortful heuristic processing. More specifically, such studies demonstrated that, compared to individuals with high confidence, individuals with low confidence process information more systematically. This outcome occurred because less confident individuals lacked confidence in their own judgments and this lack of confidence led them to feel a need for further processing (e.g., Mackie, Asuncion, & Rosselli, 1992). Similarly, less confident individuals were more likely to engage in effortful cognitive processes such as social comparisons (Pelmham & Wachtzum, 1995).

In this study, we extend the findings of prior research and conclude that power causes individuals to feel confident, which in turn leads them to process any persuasive messages in a less effortful manner because their level of confidence drives them to process information either in an effortless, heuristic fashion or an effortful, systematic fashion (Chaiken, Liberman, & Eagly, 1989; Eagly & Chaiken, 1993). Specifically, we predict that the powerful will pay less attention to a message than the powerless, resulting in poorer recall, which is a proxy measure of attention (Murphy & Anundsen, 1981). Many conceptual models of advertising posit that “consumers

Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Selective processing based on goal</th>
<th>Selective processing based on stereotype</th>
<th>Feeling confidence</th>
<th>Abstract thinking</th>
<th>Dependent variable</th>
<th>Mediator</th>
<th>Overall feeling of confidence</th>
<th>Confidence in one’s thought</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Anderson and Galinsky (2006)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Briol et al. (2007)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Déprez and Fiske (1999)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Ebenbach and Keltner (1998)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Fiske (1993)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Fiske and Déprez (1996)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Goodwin et al. (2000)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Guinote (2007b)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Guinote (2007c)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Guinote (2008)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Magee et al. (2010)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Smith et al. (2008)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Smith and Trope (2006)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Weick and Guinote (2010)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>
must attend to the ad as a necessary first step; that is, they must devote processing capacity to it” (Pechmann & Stewart, 1990: 189; see also MacInnis & Jaworski, 1989). Specifically, Pechmann and Stewart (1990) demonstrated that attention was a critical mediating variable influencing the effectiveness of advertising. They proposed that a more promising approach might be to motivate consumers to pay closer attention to advertisements by ensuring that the ad is personally relevant; i.e., that it is instrumental in achieving their personal goals and values. The present study takes this view. Moreover, we focus on memory, based on previous studies that show significant positive relationships between awareness of top-of-mind brand names and measures such as purchase behaviors and market shares (Holman & Hecker, 1983). Finally, we expect that the powerful will be less persuaded by a message than the powerless because attention indicates weight in a judgment (Fiske, 1980) and in building or changing one's attitude (Lavidge & Steiner, 1961). With respect to the mediating effect of confidence, this paper differs from Briñol et al. (2007) in several ways. First, to explore the relationship between power and confidence, we induce power prior to a message, not after. Briñol et al. (2007) induced power prior to a message in their second and fifth studies. However, they did not show how confidence was associated with the effect of power. Rather, they established a mediating analysis in Study 4, although they induced power after the message in this study. Second, we highlight the mediating role of confidence that is not relevant to a message, whereas Briñol et al. (2007) examined the mediating effect of confidence in a message-related thought. Third, we investigate whether goal-relevance mediates the effect of power. Finally, our target variables are attention, recall, and persuasion, while Briñol et al. (2007) focused on attitude change.

2.3. Power and goal pursuit

Selective information processing exists in situ and occurs when an individual selectively focuses on expectation-consistent evidence while neglecting expectation-inconsistent evidence (Sanbonmatsu, Posavac, Kardes, & Mantel, 1998). Because cognitive capacity has limits, individuals select primary factors that have greater relevance to accessible constructs, such as needs, goals, expectations, etc. (Higgins, 1996). Several studies on power have demonstrated that the powerful act in a greater number of variable ways (Galinsky et al., 2003). More specifically, Overbeck and Park (2001) reported that the powerful used more flexible information-processing strategies that were in line with task demands. According to this study, when information was relevant to the task at hand and stereotypes were irrelevant, the powerful engaged more in the processing of individuals' information about other people. The authors concluded that power increased or decreased one's cognitive effort in processing, depending on whether the given information was relevant to one's accessible constructs.

In terms of goal pursuit, recent research suggested that the more power individuals possess, the more focused they are on their own goals (Van Dijke & Poppe, 2006) because power increases the ability to process information selectively and fosters attentional flexibility based on goal-relevance (Guinote, 2007a, 2008). Consistently, the Situated Focus Theory of Power demonstrated that the powerful focus more narrowly on their goals, act in more goal-consistent ways, and pay closer attention to task-relevant information compared to the powerless because “power promotes a goal-directed attentional focus” (Guinote, 2010: 595). In contrast, the theory argued that the presence of information irrelevant to their current goals fails to influence the powerful. This is consistent with the concept of situated cognition, which argues that cognition is situated in response to environmental cues and is adaptive to the perceiver's current goals (e.g., Smith & Semin, 2007).

Based on these concepts, this study proposes that, upon the activation of the participants' goals, the goal-relatedness of any subsequent information will moderate the effect of power on cognitive processing. Due to the way that power increases goal-directed activity (Weick & Guinote, 2010), we predict that the powerful will be more persuaded by a message than the powerless when that message is relevant to their situational goal, regardless of having confidence enhanced by power. However, we predict that the reverse effect would occur in a neutral situation, where neither preexisting goals related to the situation nor active situational goals exist. Specifically, we predict that, in this kind of situation, the powerful will invest in less effort in processing a persuasive message and be less persuaded by it than the powerless.

To obtain more direct evidence, we assessed attention, memory, and persuasion as our target variables. As shown in Table 1, previous studies examined the effect of power on attention and persuasion. However, there has been little research exploring whether such effects come from a systematic or heuristic modes of processing. If the powerful do indeed process a message systematically, they should exhibit higher recall of the message. In contrast, if heuristic processing was the primary mode of attention and persuasion, the powerful should not reveal any advantages in their recall. To investigate the type of processing underlying the power effect, we test the effects of power on recall. If power induces less elaborate processing, we should then find evidence that participants without power recall information better than those with power.

3. Overview of the current research

Our basic hypothesis is that perceived power and cognitive processing are intimately associated, such that activating the concept of power should decrease the tendency to process further information effortlessly. In addition, we propose that the effect of power varies depending on a participant's motivation to pursue a goal. To test our predictions, we used multiple power manipulations. Prior research demonstrated that exposing individuals to semantic cues related to power (e.g., Bargh et al., 1995) or to recollections of experiences with power (Galinsky et al., 2003) could activate the psychological properties of power. Moreover, such research showed that priming individuals by making them act (Schubert, 2005) or think (Galinsky et al., 2003) in power-related ways or exposing them to power-related semantic cues (Chen et al., 2001) affected the individuals in a manner similar to the effect of having actual control over outcomes. This effect could create different levels of perceived power. The present research manipulated perceived power by placing participants in a hierarchical structure (Experiment 1) or by priming participants, i.e., by asking them to recall an event in which either they had power over another person or someone else had power over them (Experiments 2 & 3).

4. Experiment 1

The first experiment examined whether power affected the allocation of cognitive resources. Specifically, we tested the hypothesis that feeling powerful results in greater confidence in an individual's mind, which in turn leads the individual to have less motivation to engage in the effortful processing of subsequent information. To elicit the perception of power, we used a role-playing game (boss vs. subordinate) that has been employed in previous studies (Anderson & Berdahl, 2002; Briñol et al., 2007; Galinsky et al., 2003).

4.1. Method

4.1.1. Participants

Participants were 70 undergraduate students (35 high-power individuals; 43 females) from a Midwestern U.S. university who
participated in the study in exchange for a $5 payment. Participants were randomly assigned to the two experimental conditions (high and low power). We excluded two participants in the high-power condition from the analysis because they reported being suspicious of the procedural deception during debriefing.

4.1.2. Procedure
Participants entered the lab in pairs and were informed that they would be participating in two unrelated studies. As previously described, one study related to social structures (but actually focused on power manipulation), while the other evaluated the effectiveness of an advertisement. First, participants engaged in a role-playing activity for the power manipulation task and were then guided to sit at desks and look at an advertisement for a bottle of mineral water on the computer screen. The advertisement contained the brand name, “Pure,” and six statements describing the target product, such as “Eco-friendly bottled water” (see Appendix A). We used the Qualtrics survey program to present all materials on the computer. Based on MacInnis and Jaworski’s (1989) method of attention measurement, the Qualtrics program unobtrusively recorded the amount of time participants looked at each page of the target advertisement, which served as a proxy for attention. Next, participants took part in a five-minute filler task, which reduced the demand effect. We then asked each participant to recall the brand name and advertising claims mentioned in the advertisement, cuing the brand name recall measure with the product category (Singh, Rothschild, & Churchill, 1988). We used the item “Please write down the brand name of the mineral water advertised on the screen,” which was coded dichotomously as 1 (complete recall) and 0 (otherwise). We also cued the claim recall measure using the product category. The item prompt was, “Please write down what was said or claimed about the advertised mineral water in the advertisement.” We coded claim recall as correct (1), partially correct (.5), or incorrect (0) for each of the six statements. These scores were summed up to calculate the overall score for claim recall, which ranged from 0 to 6. Next, participants completed a questionnaire, which included a general measure of confidence. Based on the procedure of Briñol et al. (2007), we asked participants to rate their own general confidence level on a 7-point scale (1 = “not confident at all” to 7 = “extremely confident”) and complete several filler items. Afterwards, the participants completed a manipulation-check item that asked them to rate how powerful they felt during the role-playing game, using a 7-point scale (1 = “not powerful at all” to 7 = “extremely powerful”). Next, consistent with Briñol et al. (2007), we asked participants to report how they felt (happy vs. sad, unpleasant vs. pleasant, depressed vs. uplifted, excited vs. relaxed) using a 7-point scale (1 = “not at all” to 7 = “very much”; α = .86) to rule out possible alternative interpretations for the association between manipulated power and mood. Finally, participants were debriefed and thanked for their participation.

4.1.3. Power manipulation
We adapted the procedure for manipulating positions of power directly from Galinsky et al. (2003, Exp. 1). Initially, we told the paired participants that they would be performing a coordination task requiring one to be the manager and the other to be the subordinate. We randomly assigned the manager and subordinate roles before the participants arrived. The managers received instructions emphasizing that they would have complete control over the work process, the evaluation of subordinates, and the division of rewards. The subordinates received instructions emphasizing that they would have no control over how the work was performed, the evaluation process, or the division of resources. Participants played this role-playing game for 5 min (for instructions, see Appendix B). We differentiated the vertical position of each role to enhance the power manipulation by asking the “manager” participant to sit on a taller and better-looking chair.

Previous studies demonstrated that this kind of role-playing game could effectively induce differential levels of perceived power (e.g., Galinsky et al., 2003; Rucker, Dubois, & Galinsky, 2011).

4.2. Results

4.2.1. Manipulation checks and hypotheses testing
The managers experienced more power (M = 6.09, SD = .58) than the subordinates (M = 2.54, SD = .70; t(66) = 22.82, p < .001), indicating that the power manipulation was effective. The results showed that power reduced the attention given to the target advertisement. Participants assigned to the high-power condition looked at the target advertisement less (M = 3.76 sec, SD = .97) than those assigned to the low-power condition (M = 5.86 sec, SD = 1.14; t(66) = −8.15, p < .001). As we expected, a significant power effect emerged in the recall of both the brand name and the advertisement’s claims. More participants in the low-power condition (82.3%) than in the high-power condition (51.52%; z = 2.43, p < .01) correctly recalled the brand name. Two independent judges who were unaware of the participants’ conditions coded the recalled claims. Interrater agreement was high (r = .96); therefore, the mean was used in our analyses. T-test results showed that power significantly affected the recall of advertisement claims, such that participants in the high-power condition had lower overall recall of the claims (M = 2.27, SD = .99) than participants in the low-power condition (M = 4.63, SD = .73; t(66) = −11.24, p < .001).

4.2.2. Mediation analyses
To test the process underlying the effect of power on attention and recall, we performed 1000 bootstrap resamples using Preacher and Hayes’s (2008) SPSS macro, as recommended by Zhao, Lynch, and Chen (2010). First, we directly tested the significance of the indirect effect (i.e., the path through the mediator) by constructing a bias-corrected 95% confidence interval (CI) around the indirect effect, where mediation occurs if zero falls outside of the interval range. The indirect effect of power on attention via confidence was significant, with a point estimate (PE) of −.86 and the interval ranging from −1.26 to 0.49, providing statistical evidence of successful mediation. Subsequent bootstrapping procedures confirmed that power had a significant indirect effect on brand name recall through confidence (PE = −.259, CI = −.436 to −1.18). Similarly, the indirect effect of power on claim recall was significant, with a PE of −1.24 and 95% confidence interval excluding zero (−2.28 to −1.19). The regression procedures also confirmed the significant mediating effect of confidence (for results, see Table 2).

Power had no effect on mood (p > .30). In Experiments 1, 2, and 3, we first conducted the analyses by using gender as an independent variable; no interaction effects of gender were found. Therefore, in the results of our simpler analyses, we will not refer to gender.

4.3. Discussion

In Experiment 1, we obtained several noteworthy findings. First, power negatively affected not only attention but also recall. Second, confidence, elicited by power, reduced the need for processing subsequent information in an effortful manner. Third, confidence mediated the negative effect of power on attention and recall. Fourth, power had an effect on processing even when it was functionally irrelevant to the persuasive message. However, one possible limitation regarding the manipulation of power remained. In Experiment 1, power manipulation required the powerful participants to direct subordinates during a task, control their access to resources, and evaluate their performances. Thus, those with power might have been more “cognitively taxed with mental planning” (Galinsky et al., 2003: 456) than those without power. Therefore, this type of manipulation might contain an inherent confound. A greater cognitive load, elevated by power, could
Table 2
Mediation model for the effect of power through confidence (Experiments 1 & 2).

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>( \beta )</th>
<th>( p )</th>
<th>Bootstrap analysisa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Indirect effect</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>Power</td>
<td>1.56</td>
<td>.00</td>
<td>.55</td>
</tr>
<tr>
<td>Brand name recall</td>
<td>Power</td>
<td>−1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conf. Power</td>
<td>−1.15</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power, conf.</td>
<td>−.82</td>
<td>.00</td>
<td>−.39</td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>−.63</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Claim recall</td>
<td>Power</td>
<td>−1.18</td>
<td>.00</td>
<td>−.50</td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>−1.52</td>
<td>.00</td>
<td>−.97</td>
</tr>
<tr>
<td></td>
<td>Power, conf.</td>
<td>−1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>−1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>Power</td>
<td>1.18</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Brand name recall</td>
<td>Power</td>
<td>−1.15</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conf. Power</td>
<td>−.82</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power, conf.</td>
<td>−.39</td>
<td>.00</td>
<td>−.75</td>
</tr>
<tr>
<td>Claim recall</td>
<td>Power</td>
<td>−1.18</td>
<td>.00</td>
<td>−.50</td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>−1.52</td>
<td>.00</td>
<td>−.97</td>
</tr>
<tr>
<td></td>
<td>Power, conf.</td>
<td>−1.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confidence</td>
<td>−1.00</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

a Both are two-sided tests.

5.1. Method

5.1.1. Participants
Seventy-six participants (38 high-power individuals; 31 females) from a midwestern American university took part in this experiment in exchange for a payment of $3. They were randomly assigned to the treatment conditions.

5.2. Results

5.2.1. Manipulation checks and hypotheses testing
Participants described themselves as having more power in the high-power condition than they did in the low-power condition (\( M_{\text{high}} = 5.71, SD = .80 \) vs. \( M_{\text{low}} = 2.82, SD = .77 \); \( t(74) = 16.09, p < .001 \)), which indicates that the power priming effort was successful. The results of a \( t \)-test replicated our earlier experiment, showing that power had a negative effect on attention. More specifically, high-power participants looked at the target advertisement for a shorter time than low-power participants (\( M_{\text{high}} = 3.92, SD = 1.15 \) vs. \( M_{\text{low}} = 6.21, SD = 1.71 \); \( t(74) = −6.85, p < .001 \)). As expected, a significant main effect of power emerged in the recall of both brand name and claims. The high-power participants (36.84%) correctly recalled the brand name less often than the low-power participants (81.58%; \( z = 3.74, p < .001 \)). Similar results were shown in the recall of the claims (\( M_{\text{high}} = 2.42, SD = 1.16 \) vs. \( M_{\text{low}} = 4.67, SD = 1.10 \); \( t(74) = 8.64, p < .001 \)).

5.2.2. Mediation analyses
To investigate the mediating effect of confidence, we conducted 1000 bootstrap resamples based on Preacher and Hayes’s (2008) and regression procedures. The results of a 95% confidence interval indicated that the indirect effect of power on attention was significantly separated from zero (−1.11 to −.39), with a \( PE = −.75 \). Similarly, bootstrapping analyses confirmed that there was a significant indirect effect of power on brand name recall (\( PE = −1.82, CI = −3.46 \) to −.29), as well as claim recall (\( PE = −1.18, CI = −1.45 \) to −.93). As displayed in Table 2, the regression procedures also revealed that the effect of power on attention and recall was mediated by confidence. Power had no significant difference on mood (\( p > .10 \)).

5.3. Discussion

Experiment 2 replicated and extended the results of Experiment 1. The results revealed the robustness of the effects that we observed in
Experiment 1 by showing how primed power affected cognitive responses to a persuasive message in a manner similar to that of manipulated power. Moreover, the findings showed that confidence also mediated the effect of primed power.

Social psychologists have demonstrated that power leads one to devote more attention to accessible constructs (Higgins, 1996). Because power promotes a goal-directed attentional focus and leads one to act in a goal-consistent manner (Weick & Guinote, 2010), people with high power have a greater tendency to pursue their current goal when the goal is situated (Slabu & Guinote, 2010), and their attentional focus can vary as a function of the accessibility of their goal (Guinote, 2007b). Based on these notions, we conducted a final experiment to examine whether the powerful processed a persuasive message more effortlessly than the powerless when the message was relevant to their current goal. We predicted that, in the goal-relevant condition, the powerful would pay greater attention to a given message and would recall the message more accurately than the powerless. However, as in Experiments 1 and 2, we predicted the opposite performance pattern when there were no pre-existing or active specific goals.

A large body of consumer research has suggested that, when consumers are willing to devote considerable effort to the processing of the central messages behind communicated information, they are likely to form attitudes based on their beliefs (e.g., Hoyer & MacInnis, 2008). Conversely, when their level of processing effort is low, consumers are passive recipients of the message and form attitudes toward the communicated product without realizing why they did so. As mentioned in our previous review, the power literature has posited that the powerful are “lazy information processors” (Guinote, 2007a: 257). Moreover, the research has shown that power reduces the tendency to devote one’s efforts to information processing, which in turn leads to a smaller differentiation between strong and weak arguments (Briñol et al., 2007) and poorer comprehension of others’ thoughts (Galinsky et al., 2006). Hence, we predicted that a persuasive message would have less effect on the attitudes of the powerful than of the powerless in a neutral setting. However, we also predicted different power outcomes with the activation of specific goals; e.g., the powerful would pay greater attention to the goal-relevant message and be more persuaded by the message if the message benefitted their goal compared to the powerless because power promotes goal-directed processing and behavior. In sum, we predicted that the interaction between power and condition (goal-situated vs. control) would determine not only attention and recall, but also persuasion, in a manner similar to the way in which attention indicates not only the beginning of information processing (Fiske, 1993) but also its weight in a judgment (Fiske, 1980).

The findings from Experiments 1 and 2 showed that, in the activated power condition, confidence, facilitated by power, led directly to poorer message-processing performance. However, these two experiments did not consider situated goals. Thus, it remained unclear whether, with the presence of a specific goal and the perception of the goal-relevance of a message, confidence would still mediate the effect of power on message processing. In addition, it was unclear whether the effect of power in the two previous experiments was due to high power or low power or both. For instance, high power might have increased confidence, but low power might not have reduced confidence. On the other hand, high power might not have increased confidence, but low power might have reduced confidence. Moreover, both situations are possible. Without controlling for the differential levels of power feelings, it was impossible to determine what was actually going on. To address this issue, we conducted Experiment 3, which included a neutral group.

6. Experiment 3

In Experiment 3, we explored the interactive effect of power and the existence of a relevant goal on the cognitive processing of a persuasive message and the mediating effect of confidence on this interaction. We first manipulated the goal-relevant or goal-irrelevant context by showing participants a series of pictures before presenting those pictures with the advertisement for mineral water (the one used in Experiments 1 and 2). Specifically, we used images associated with actions. According to embodiment theorists, merely thinking about a behavior facilitates the mental simulation of and the tendency to engage in that behavior (e.g., Chartrand & Bargh, 1999). This suggests that, when individuals perceived information implying certain actions or emotions, they are likely to adopt an active role that mentally simulates the implied actions or emotions (Jeannerod, 1994; Prinz, 1997). In Experiment 3, we took this view and manipulated goal relevance by priming participants for a goal associated with action. Lee and Min (2010) manipulated the association of a goal with an action as follows: one group was presented with images of people running a marathon in order to induce them to mentally simulate exercising; images of gardens were presented to the other group. The authors found that those participants who had viewed images of marathon runners felt more tired and thus, performed worse on a handgrip task than those participants who had viewed images of gardens. That is, the images of marathon runners motivated the participants to relive their fatigue. Conceptually consistent with Lee and Min’s (2010) priming of action-associated goals, we manipulated two conditions (related/unrelated to thirst-quenching goals) by using images of marathon runners (the goal-situated condition) and lamps (the control condition). With the images of marathon runners, we expected that participants would feel thirstier and would thus activate a thirst-quenching goal. However, with images of lamps, we expected that participants would not have any specific thirst-quenching goal. Therefore, with the subsequent perception of a message about mineral water, we expected that those who had been primed with marathon images would make more effort to process the message and would have more favorable attitudes toward the target product than those who had viewed images of lamps.

6.1. Method

6.1.1. Pretest and development of stimuli

Following the procedure developed by Lee and Min (2010), we presented 13 participants with pictures of six products and asked them to rate the relevance of these products for people who run marathons, using a 7-point scale (1 = “very disadvantageous” to 7 = “very advantageous”). Scheffé’s homogeneous subset test (F(5, 72) = 429.33, p < .05) revealed that the products fell into three categories with the following relevancies to a running goal: advantageous (ice pack, water, orange; M = 6.56), disadvantageous (soup, coffee; M = 1.19), and neutral (pencil; M = 3.92). In the second pretest, we presented 39 participants with 20 images of marathon runners (N = 29) or lamps (N = 19). We then exposed them to images of six products (ice pack, water, soup, coffee, and pencil) and asked them to rate each product on a 7-point scale (1 = “very undesirable”; 7 = “very desirable”). In addition, we asked participants to rate the relevance of the products to the images in order to rule out possible alternative interpretations of associations between the images and the product, which were categorized as advantageous, disadvantageous, and neutral with respect to running a marathon. To analyze the data, we first averaged the participants’ evaluations of advantageous products (α = .92) and disadvantageous products (α = .75) to create two indices. We then conducted a 2 (context: marathon runners vs. lamps) × 3 (product: advantageous vs. neutral vs. disadvantageous) repeated measures ANOVA with the product as a within-participant factor on the products’ evaluation. The results indicated that both the main effect of the context (F(1, 37) = 1.57, p > .10) and the product were not significant (F < 1). However, the context × product interaction was significant (F(1, 37) = 107.22,

interaction was significant (α = .88) and the marathon-disadvantageous products were highly intercorrelated (α = .79). To analyze the data, we averaged the participants’ ratings for relevant and irrelevant categories. We then conducted a 2 (context) × 3 (product) repeated measures ANOVA by using the product as a within-participant factor for perceived relevance. The results indicated that both the main effect of the product and that of the condition were not significant (Fs < 1), whereas the context × product interaction was significant (F(1, 37) = 9.14, p < .01). Planned contrasts revealed that participants who viewed images of lamps showed no difference in their ratings for the three types of products (Madv = 2.94, SD = .59 vs. Mneutral = 3.11, SD = .71 vs. Mdisadv = 2.82, SD = .61; F(1, 18) = 1.78, p > .1). However, those who viewed images of marathon runners rated the marathon-advantageous products as being more relevant to the images than the neutral product (Madv = 5.72, SD = .54 vs. Mneutral = 3.65, SD = .67; F(1, 19) = 141.54, p < .001) and the neutral product as being more relevant than the marathon-disadvantageous products (Madv = 2.78, SD = .53; F(1, 19) = 24.37, p < .001). Therefore, we used the images of marathon runners and lamps to prime goals that were related or unrelated to the target product (bottled mineral water).

6.1.2. Participants and procedure
Participants were 184 undergraduates (63 powerful individuals vs. 64 powerless individuals; 82 females) at a university in Korea. All participants received monetary compensation for participating. We randomly assigned the participants to their treatment condition. The study consisted of two phases: power manipulation and goal priming. First, both the powerful and powerless participants engaged in a power manipulation task. As in Experiment 2, participants described an incident in which they had power over another person (high-power condition) or another person had power over them (low-power condition). There was no power manipulation for the neutral participants. Power was coded as high (1), neutral (0), and low (−1). After they completed this task, we asked participants to rate how powerful they felt and to report their confidence and mood on the same scales used in Experiments 1 and 2. Upon completion of this task, we presented them with images of people running a marathon (goal-situated condition) or lamps (control condition). Participants in the goal-situated condition (N = 98) first read, “You will be presented with some pictures of marathon races. Please imagine that you have always wanted to run a marathon. You participate in marathons.” We told participants in the control condition (N = 86) that they would be seeing some pictures of lamps. After all participants viewed the appropriate images, we exposed them to the target advertisement for the eco-friendly mineral water employed in the previous two experiments, and measured their attention by the amount of time they spent looking at the target advertisement. Participants then completed a 10-minute filler task. Next, we asked them to recall the brand name and claims. They then rated the advertised mineral water using the four 7-point scale items anchored by “not at all (1)” and “very much (7)” (favorable, like, good, and positive), which was employed by Lee and Min (2010). The ratings were highly intercorrelated (α = .89); we averaged them to create a composite index of attitudes. Finally, all participants indicated their state of mood and answered questions related to their interest in marathons and the frequency with which they participated in marathons.

6.2. Results
6.2.1. Manipulation checks and control measures
The results of a 3 (power: high vs. low vs. neutral) × 2 (condition: goal-situated vs. control) ANOVA with powerful feeling as a between-participant factor indicated that manipulated power had a significant effect on powerful feeling (F(2, 178) = 61.01, p < .001). Subsequent contrasts revealed that participants in the high-power condition felt more powerful (M = 5.97, SD = .67) than those in the neutral condition (M = 3.72, SD = .45; t(178) = 9.64, p < .001). In turn, participants in the neutral condition felt more powerful than those in the low power condition (M = 2.63, SD = .76; t(178) = 4.64, p < .001), implying that power priming was successful. A 3 (power) × 2 (condition) ANOVA, including control measures as covariates, revealed that mood, interest in marathon running, and frequency of participating in marathon races did not produce any significant effects (all ps > .13).

6.2.2. Attention, claim recall, and attitudes
As displayed in Table 3, the results of a 3 (power) × 2 (condition) multivariate analysis of variance (MANOVA) with attention, claim recall, and attitudes as the dependent variables of the analysis indicated significant main effects of power and condition. Subsequent contrasts revealed that high-power participants showed better performance in attention and claim recall and indicated more favorable attitudes than low-power participants, who in turn were better in attention and claim recall and were more favorable than neutral participants. Participants in the goal-situated condition were more likely to pay attention, recall claims, and be more favorable than those in the control condition. More importantly, the MANOVA results revealed a significant interaction effect of power and condition on the dependent variables. Similar to the results of Experiments 1 and 2, planned contrasts showed that high-power participants in the control condition paid less attention, recalled claims less, and were less favorable than neutral participants, who in turn performed worse in attention and claim recall and were less favorable than low-power participants. However, as predicted, high-power participants in the goal-situated condition showed better performance in attention and claim recall and were more favorable than low-power participants, who in turn were better in attention and claim recall and were more favorable than neutral participants (for results, see Web Appendix).

6.2.3. Brand name recall
Next, we conducted a 3 (power) × 2 (condition) logistic regression for the brand name recall to test our hypotheses (for results, see Web Appendix). The results indicated that power had no significant main effect (β = −.18, SE = .21, p > .50). However, the main effect of condition was significant (β = 2.24, SE = .49, p < .001), showing that participants better recalled the brand name when they were exposed to images of marathon runners instead of lamps (Mgoal = 74%, SD = .44 vs. Mcontrol = 41%; SD = .94; z = 4.78, p < .001). More importantly, the power × condition interaction had a significant effect (β = −.74, SD = .21, p < .001). For a better understanding, we conducted z-tests for proportion pairs. We first divided participants into the goal-situated and control groups and compared the accuracy of brand name recall between high-power, low-power, and neutral participants in each condition. With images of marathon runners, high-power participants recalled better than low-power participants (Mhigh = 91%, SD = .29 vs. Mlow = 71%, SD = .46; z = 2.20, p < .05), who in turn recalled better than neutral participants (Mneutral = 51%, SD = .50; z = 1.72, p < .10). With
images of lamps, however, high-power participants recalled worse than neutral participants (M_{high} = 14\%, SD = .41 vs. M_{neutral} = 33\%, SD = .51; z = 1.73, p < .10), who in turn recalled worse than low-power participants (M_{low} = 55\%, SD = .51; z = 1.72, p < .10).

### 6.2.4. Multiple mediation analyses

Previous research has demonstrated that goal compatibility facilitates attention, and therefore, more elaborate processing occurs when information is compatible or incompatible with the individuals' goal (e.g., Higgins, 1996). Specifically, Aaker and Lee (2001, Exp. 2) showed that individuals are more likely to recall goal-compatible condition information than goal-incompatible information because goal compatibility facilitates elaboration. Haugtvedt and Petty (1992) found that, with increased attention, individuals in a goal-compatible condition paid more attention to the information, and thus, recall and persuasion were less likely for high-power participants than for low-power and neutral participants. Hence, for goal-compatible information, the need for goal pursuit, driven by power, induced participants to pay more attention to the information, and thus, recall and persuasion were more likely for high-power participants than for low-power and neutral participants. Hence, we concluded that the effect of power and the role of confidence as a mediator of the power effect on cognitive responses varied according to the type of information (i.e., goal-compatible vs. goal-incompatible). This is consistent with the notion that "powerful individuals process more extensively information that is relevant to accessible constructs, compared to information that is irrelevant to these constructs. ... As a consequence, their attentional focus can vary as a function of the constructs that are accessible in a given setting" (Guinote, 2007b: 685).

### 6.3. Discussion

As expected, there were significant differences in the effects of power on processing between the conditions. High-power participants performed best in the goal-situated condition and worst in the control condition. In addition, the performances of high-power participants were better than those of low-power participants in the goal-situated condition and worse in the control condition. Furthermore, low-power participants' performances were more positive than those of the neutral participants across both conditions. These results indicate that feeling powerful facilitated goal pursuit when a goal was situated and thus, high-power participants were more likely to engage in the effortful processing of goal-relevant information.

---

**Table 3**

Multivariate analysis of variance for attention, claim recall, and attitudes* (Experiment 3).

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>High</td>
<td>5.74</td>
<td>.29</td>
<td>37.53</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.58</td>
<td>.82</td>
<td>High–low: 3.90</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>5.00</td>
<td>1.26</td>
<td>Neutral–low: 6.92</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Goal-sit.</td>
<td>5.89</td>
<td>2.47</td>
<td>133.03</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.62</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power × Condition</td>
<td>Goal-sit.</td>
<td>8.41</td>
<td>1.07</td>
<td>High–low: 10.44</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.76</td>
<td>1.02</td>
<td>Neutral–low: 4.49</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>5.20</td>
<td>1.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.69</td>
<td>1.00</td>
<td>High–neutral: 2.07</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.39</td>
<td>.50</td>
<td>Neutral–low: 4.05</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.76</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Claim recall</strong></td>
<td>Power</td>
<td>4.21</td>
<td>1.58</td>
<td>9.68</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.67</td>
<td>.72</td>
<td>High–low: 1.6</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4.23</td>
<td>.83</td>
<td>High–neutral: 4.18</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Goal-sit.</td>
<td>4.67</td>
<td>1.91</td>
<td>152.03</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.34</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power × Condition</td>
<td>Goal-sit.</td>
<td>5.54</td>
<td>1.40</td>
<td>High–low: 6.17</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.76</td>
<td>1.02</td>
<td>Neutral–low: 4.49</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4.49</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.66</td>
<td>.81</td>
<td>High–neutral: 4.27</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.46</td>
<td>.49</td>
<td>Neutral–low: 2.40</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.91</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attitudes</strong></td>
<td>Power</td>
<td>4.44</td>
<td>1.77</td>
<td>12.40</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>3.77</td>
<td>.71</td>
<td>High–low: 6.4</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4.36</td>
<td>.80</td>
<td>Neutral–low: 4.63</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Goal-sit.</td>
<td>4.91</td>
<td>1.07</td>
<td>201.94</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.40</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power × Condition</td>
<td>Goal-sit.</td>
<td>5.93</td>
<td>.50</td>
<td>High–low: 7.77</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>4.05</td>
<td>.79</td>
<td>Neutral–low: 3.24</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4.62</td>
<td>.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.68</td>
<td>.84</td>
<td>High–neutral: 4.31</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.48</td>
<td>.50</td>
<td>Neutral–low: 3.02</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.04</td>
<td>.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Two-sided tests.
than low-power participants. In contrast, low-power participants were more likely to engage in effortless processing than neutral participants. Hence, the results support the notions that “the powerful, relative to the powerless, have greater ability to focus attention in line with the demand of the task” (Guinote, 2007b: 685) and that “when individuals feel powerless, they should devote more attention to others” (Keltner, Gruenfeld, & Anderson, 2003: 276).

In Experiment 3, we obtained several noteworthy findings. First, by replicating previous research, we found that, with a specific situated goal, power facilitated goal pursuit, which in turn increased participants’ cognitive processing of the goal-related message. However, with no such goal, power had a negative effect on the processing of subsequent information. Second, we found that confidence mediated the effect of power not only on attention and memory, but also on persuasion. However, the mediating effect of confidence was not significant when power facilitated goal pursuit. These results indicate that, with a specific situated goal, power positively affects goal pursuit and confidence has no influence on the relationship between power and goal pursuit, although power induces confidence. Finally, bootstrapping tests for multiple mediators revealed that, when participants processed goal-irrelevant information, confidence mediated the effect of power on memory and attitudes, whereas attention was not a mediator. In sum, our results suggest that power has differential (or opposite) effects on cognitive processing, depending on the presence of relevant goals. Furthermore, the mediating effect of confidence on the relationship between power and cognitive response can weaken when one pursues a goal. Thus, when further information is relevant to a pre-existing goal, power may directly influence the amount of processed information. However, when further information is not relevant to the pre-existing goal, power may influence processing through confidence.

7. General discussion

Current research provides convergent evidence that power influences cognitive processing by affecting one’s level of goal pursuit. In a neutral setting, high-power participants are less likely to pay attention to the persuasive message and less likely to recall it compared to low-power participants (Experiments 1 and 2). However, with a message-relevant goal (Experiment 3), high-power participants are more likely to process goal-relevant information than low-power and neutral participants. In particular, when there is no situated goal, confidence mediates the effect of power across the three experiments, whereas when a relevant goal is situated prior to the message, the mediating effect of confidence is not found. Hence, our results generalize prior findings, suggesting that power fosters attentional flexibility (Guinote, 2007b) and goal pursuit (Guinote, 2007c) by showing that the interaction of power and goal pursuit affects the extent to which one engages in effortful processing. Moreover, we extend the notion of Briñol et al. (2007) that confidence in one’s thoughts mediates the effect of power after a message by finding that overall feelings of confidence mediate the effect of power prior to a message and the mediating effect only occurs when a message-relevant goal is not situated.

The results provide a rich avenue for future research on the relationship between power and emotions. Previous studies have demonstrated that increased power triggers positive emotions, whereas reduced power triggers negative emotions (for review, see Keltner et al., 2003). In our experiments, however, power had no effect on mood, which is consistent with the findings of previous studies that manipulate power by placing participants in a hierarchical structure (Briñol et al., 2007), allowing them to control real resources (Anderson & Berdahl, 2002), and having them describe experiences of their own in which they felt powerful or powerless (Galinsky et al., 2003). The different sources of power may explain this inconsistency. Anderson and Berdahl (2002) investigate the effects of power on emotional valence and show that personality dominance fosters positive emotions and limits negative emotions, whereas power manipulation does not affect emotional experience. They argue that the sense of power relates more to emotions than to manipulated power. Because emotions can vary “from situation to situation in terms of their intensity and appropriateness” (Tiedens & Linton, 2001: 986), an examination of the relationship between emotional experiences and power, in terms of different sources of power, should be meaningful.

Furthermore, we note that the mediating role of attention and confidence in the effect of power on memory and persuasion are inconsistent across goal-relevant and goal-irrelevant conditions. Specifically, if information is not goal-relevant, confidence mediates the effect of power on memory and persuasion, but attention does not. How can confidence affect memory and persuasion, if not through altered attention? Additional research is needed to examine how and why confidence, induced by power, affects memory and persuasion directly without the effect of attention.

Personal power is a related issue. Previous studies (e.g., Van Dijke & Poppe, 2006) have suggested that social power is the possibility of influencing others whereas personal power is the ability to do whatever one wants to do and freely make one’s own decisions without the influence of others. Lammers, Stoker, and Stapel (2009) demonstrate that these two types of power are inversely associated with independence and interdependence, and whether their effects are opposing or parallel to one another depends on what individuals do. Specifically, they argue that social power and personal power have opposite effects on stereotyping but have similar effects on behavioral approaches. In the present study, we find consistent effects of power on attention, memory, and persuasion. However, the manipulated power in our study is more related to social power than to personal power. In this regard, future research should compare the effects of personal power on attention, memory, and persuasion to those of social power. In addition, future research should examine the influence of power arising from various sources, such as company reputation, financial resources, and country of origin. In this study, we manipulated differential levels of power using role manipulation and experiential priming of power, which is a common method in social psychology (Galinsky et al., 2003, 2008; Guinote, 2010). However, in the context of marketing, there is a gap between the actual experience of power and a retrospective essay about a power-related
experience. Thus, to increase the external validity of these findings, future research should explore the effects of power in real-world settings.

Previous studies on social power have suggested that power promotes a relatively effortless, automatic, and associated use of cognitive heuristics (Chaiken et al., 1989) and causes individuals to judge others unsystematically (Mullen, Brown, & Smith, 1992). However, Miyamoto and Ji (2011) demonstrate that power promotes context-independent analytical processing. Their findings reveal that, when individuals have the capacity to influence others, the goals directed toward others become more salient than the environmental context, making them more likely to engage in the analytical processing of focal objects than in the holistic processing of the surrounding context. The present study sheds some light on what determines the amount of processing effort for the powerful and suggests that the goal compatibility of information influences whether power facilitates more or less effortful processing. Thus, another avenue for future research is the exploration of the relationships among power, goal compatibility, and the type of cognitive processing.

Our results have important implications for advertisers and retailers. Across a wide range of marketing environments, marketers often use strategies to make customers feel powerful by enhancing their subjective stature. Firms frequently use phrases such as “the customer is always right” and “the customer is king” to convey their eagerness to put their customers first. By attempting to live up to such phrases, firms strive to make their customers feel special. However, such strategies do not always lead to customer satisfaction. Former Continental Airlines CEO Gordon Bethune argues that the concept of “the customer is always right” is not appropriate for his company and notes that the strategy of making customers feel powerful induces unreasonable and intransigent behaviors (Bethune & Huler, 1998). What leads customers to engage in such behaviors? One can simply argue that feeling powerful and special causes customers to have an increased sense of entitlement. Thus, although they may agree that what they are demanding is not the standard, they may feel powerful enough to ask for special treatment. Otherwise, enhanced power could lead customers to consider only their own objectives, such as convenience or saving money, because power facilitates the pursuit of their current goals. In contrast, power may reduce the customer’s need for processing company messages, such as common rules for using a service. This suggests that firms should augment their customers’ perceived power cautiously. Instead of “the customer is always right,” the slogan “we address what you want before you ask” may slightly shift the center of power from the customer to the firm, which in turn may motivate the customer to pay more attention to the firm’s messages.

Recently, Smith and Trope (2006) and Smith, Wigboldus, and Dijksterhuis (2008) demonstrated that the relationship between power and abstract thinking is bidirectional, implying that activating power facilitates the relevant representation of abstract thinking and that inducing individuals to think more abstractly instead of concretely can make them feel more powerful. Because power can induce individuals to pay more attention to an advertisement and be more susceptible to the message when the advertisement has a goal-compatible message, marketers should focus on the customers’ style of thinking (i.e., abstract vs. concrete) when presenting marketing messages. For example, if marketing messages are relevant to target customers’ goals, then it may be beneficial to help customers feel more powerful. In these cases, inducing customers to think about messages more concretely may be helpful. We leave this issue to future empirical research.

Appendix A. Advertisement used in experiment 1, 2, and 3

**Eco-friendly Bottled Water Pure**

- Our water comes from an internationally certified organic source
- Our bottles are
  - made from plants not crude oil
  - made with 20% less plastic than the average half-liter bottle
  - flexible so they are easier to crush for recycling
- We donate more than $10 million annually to help save the Amazon rainforest

Appendix B. Instructions for power manipulation in experiment 1

1. Instructions for manipulation of status with high power (The Manager Role):

As a manager, you are in charge of directing the subordinates at a medium-sized supermarket in selling a brand of shampoo on special offer. You will decide how to structure the selling process and the standards by which the work is evaluated. In addition, you will evaluate the subordinates at the end of the session through a private questionnaire, that is, they will never see your evaluation. The subordinates will not have an opportunity to evaluate you. Thus, as a manager, you will be in charge of directing the selling process, evaluating your subordinates, and determining the rewards you and your subordinates will receive.

2. Instructions for manipulation of status with low power (The Subordinate Role):

As a salesperson, you will have the responsibility for selling a brand of shampoo on special offer according to the instructions given to you by your manager. Your manager will call you in to give you the instructions when ready. Your manager will decide how to structure the selling process and the standards by which the work is evaluated. Which task you will complete will be decided by the manager. In addition, you will be evaluated by the manager at the end of the session. This evaluation will be private, that is, you will not see your manager’s evaluation of you. You will not have an opportunity to evaluate your manager. Only the manager will be in charge of directing the selling process, evaluating your performance, and determining the rewards you will receive.
Product development capability and marketing strategy for new durable products

Sumitro Banerjee a,⁎ , David A. Soberman b,1

a ESMT European School of Management and Technology GmbH, Schlossplatz 1, 10178 Berlin, Germany
b Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada

A R T I C L E   I N F O
Article history:
First received in 15, November 2010 and was under review for 10 months
Available online 25 April 2013
Area Editor: Eitan Muller
Keywords:
Product development
Marketing strategy
Durable goods
Quality
Signaling game

A B S T R A C T

Our objective is to understand how a firm’s product development capability (PDC) affects the launch strategy for a durable product that is sequentially improved over time in a market where consumers have heterogeneous valuations for quality. We show that firms’ launch strategies are affected by the degree to which consumers think ahead. However, only those firms’ strategies that have a high PDC are affected by the observability of quality. When consumers are myopic and quality is observable, both high and low PDC firms use price skimming and restrict first-generation sales to consumers with a high willingness to pay (WTP). A high PDC firm, however, sells the second generation broadly, while a low PDC firm only sells the second generation to consumers with a low WTP. When consumers are myopic and quality is unobservable, a firm with a high PDC signals its quality by offering a low price for the first generation, which results in broad selling. The price of the second generation is set such that only high WTP consumers will buy. A firm with a low PDC will not mimic this strategy. If a low PDC firm sells the first-generation broadly, it cannot discriminate between high and low WTP consumers. When consumers are forward-looking, a firm with a high PDC sells the first generation broadly. This phenomenon mitigates the “Coase problem” that is created by consumers thinking ahead. The high PDC firm then only sells the second-generation product to the high WTP consumers. In contrast, a firm with a low PDC does the opposite; it only sells the first generation to high WTP consumers and the second generation broadly.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The quality of a product that provides unique value to consumers is often affected by the firm’s product development capability. In some cases, firms that introduce new consumer electronics and appliances such as digital photo frames, specialized equipment, software, or high-end sporting goods have a well-known track record of introducing improved versions of their products over time. In other cases, firms are either unknown or have made limited improvements to products that have been in the market for a relatively long time. Consider the evolution of the video game console market in the 1990s. In 1994, Sony entered the market and became the market leader with Playstation (PS) followed by the even more successful Playstation 2 (PS2) in 2000. PS2 offered improved user benefits such as internet connectivity and an inclusive DVD player (see for example, Ofek, 2008). Because Sony was new to the video console market, it is possible that purchasers of the first PS may not have foreseen the launch and benefits of PS2 when making the decision to buy PS.

In contrast, the launch of successive products in some markets is more predictable. Here, it is likely that consumers account for the potential benefits of future products when making a purchase decision. For example, since the launch of Apple’s widely successful iPod in 2001, new and improved versions of this product have been introduced with high regularity, offering better storage, higher song capacity, improved screens, and increased functionality such as video and touch.2

When a firm develops new product versions by improving quality or performance over time, sales of a first (early) generation product often hinder the profitability of subsequent versions. Consumers who buy the first-generation product will have lower willingness to pay (WTP) for a new generation of the same product because they already have a functioning product. Accordingly, a supplier is restricted to charging existing consumers a maximum price that is equal to the incremental value of the new generation product. In other words, high performing, early generation products limit the price that can be charged for subsequent generations.3

⁎ Corresponding author. Tel.: +49 30 21231 1520; fax: +49 30 21231 1281.
E-mail addresses: sumitro.banerjee@esmt.org (S. Banerjee), david.soberman@rotman.utoronto.ca (D.A. Soberman).
1 Tel.: +1 416 978 5445; fax: +1 416 978 5433.

0677-8116/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijresmar.2013.01.003

3 This is also the basis for the Coase problem whereby a durable good monopolist is not able to implement time-based discrimination due to customers’ understanding that the price will drop over time (Coase, 1972).
When launching the first product, a firm therefore faces the choice of employing price skimming (i.e., charging a high price and selling to a limited number of consumers with high WTP) or price penetration (i.e., selling the product to a broad set of consumers including those with low WTP). Price penetration may restrict the ability to charge higher prices for second-generation products because most potential buyers already have a first-generation product. We use the term “target breadth” as a shorthand term to describe the marketer’s choice of price skimming or price penetration described above.

To analyze how marketing strategy (i.e., the pricing of unique products over time) is affected by a firm’s product development capability (PDC), we propose a simple model. A model with two periods is the simplest way to represent firm dynamics in a context where the first generation of a product is improved upon through product development. We restrict the firm decisions to pricing for the first and second-generation products, and the level of investment in product development at the end of the first period to improve the product for the second period. These decisions constitute the most parsimonious set of decisions that can be used to understand how marketing strategies (the pricing of products over time) are affected by PDC.

We also examine how a firm’s strategy is affected by environmental factors such as how “deeply” (i.e., how far ahead) consumers think about a purchase and also how easy it is for consumers to assess the quality of products prior to purchase. In economics, unobserved quality is an important cause of market failure and the basis for a substantial amount of literature, including Akerlof (1970). We analyze how a firm responds when it faces the problem of adverse selection, i.e., if it has a high quality product, but consumers may not be willing to pay for high quality because they cannot be sure that the product is, in fact, high quality. In particular, we examine whether a firm will simply charge a price that is based on the expected value of a product or attempt to signal its high quality to consumers through actions. The questions of the paper can be summarized as follows:

a) How does a firm’s PDC affect its introductory marketing strategy in terms of pricing (which determines “target breadth”) and its investment in product improvement when consumers are either myopic (they consider only the current benefits offered by the product) or forward-looking (they consider the expected value of the product in the future as well)?

b) How are a firm’s launch strategies affected when consumers cannot assess product quality by inspection?

The key findings of our analysis are as follows:

1. A firm with a high PDC sets prices so that it sells both first- and second-generation products to consumers with a high WTP. In contrast, a firm with a low PDC focuses on intertemporal price discrimination and sells to each consumer type only once.

2. We show that the unobservability of quality changes the strategy of a firm with a high PDC because the firm will signal its quality by implementing penetration pricing (a strategy that a firm with a low PDC finds unattractive). This strategy leads to the unexpected observation that the first generation of a high quality product is sometimes priced lower than the first generation of a low quality product.

3. We show that when consumers are forward-looking, the launch strategies of firms change independently of their PDC. The change in strategy is driven by a reduced WTP of consumers for first-generation products.

Interestingly, the market strategies that are employed by a firm with a high PDC when quality is not observable and when consumers are forward looking are identical, namely: market penetration in period 1 and selling only to high WTP consumers in period 2. However, the reasons for adopting “market penetration in period 1 and restricting sales to high WTP consumers in period 2” are very different. In the former case, it is the firm’s desire to signal its quality that leads to the lower introductory price. In the latter case, it is the consumers’ lower WTP that leads to the lower introductory price.

The remainder of the paper is organized as follows: Section 2 provides a brief literature review. In Section 3 we present a model of a monopolist selling to two types of myopic consumers. In Section 4, we first present the optimal strategies of the firm when quality is observable as well as when it is unobservable. Similarly, in Section 5, we present a model for forward-looking consumers when quality is observable as well as when it is unobservable.

2. Literature review

In many categories, new product generations appear on a regular basis. Nevertheless, research (e.g., Abernathy & Utterback, 1978) suggests that technological constraints and uncertainty inhibit the willingness of firms to introduce new product generations. With uncertainty, a sequential strategy is both “information yielding,” compared to an all-or-nothing crash program (Weitzman, Newey, & Rabin, 1981), and beneficial in a context of network externality (Ellison & Fudenberg, 2000, Padmanabhan, Rajiv, & Srivisan, 1997). These benefits, however, are balanced by the reluctance of consumers to trade up to a new generation product (with higher marginal costs) when they already possess a functional first-generation product. Another factor driving the sequential generation of products is competition, either in R&D (“R&D races”) or in markets.

On the one hand, incumbent firms may invest more than entrants in R&D for subsequent innovations due to intellectual property rights and the diffusion of new products (Banerjee & Sarvary, 2009). On the other hand, prior success in R&D allows firms to gain reputation. Firms therefore trade-off R&D investment with reputation building (Ofek & Sarvary, 2003). In the absence of intellectual property rights, the possible entry of imitators may also drive incumbents to invest in developing a higher quality of new products (Purohit, 1994). Our research examines why a firm with market power may develop new product generations in the complete absence of competitive threats. Our objective is to show how launch and targeting strategies are affected by three factors: a firm’s PDC, the observability of product quality, and the degree to which consumers think ahead when making a purchase decision in the present.

Our research is also related to the durable goods literature which is reviewed by Waldman (2003). Generally, durable goods literature focuses on the effect of secondhand markets, the role of commitment to future price (or quality), and adverse selection between new and used goods. Recently, the literature has examined the role of pricing in markets where new products are launched in a context of old (or used) products. The present work highlights a Coasian time inconsistency problem because of which the monopoly price for a current product is lower due to the expected launch of products in the future. A firm may therefore offer a lower quality current product to credibly commit to price and quality in the future (Dhebar, 1994). Similarly, Moorothy and Png (1992) show that a monopolist should launch a high quality product before launching a low quality product. The firm may, however, face difficulties in developing a high quality product first because a better performing product often requires additional R&D (Langinier, 2005). Moreover, the monopolist may want to sell a higher quality product later if it does not discriminate between past and new buyers (Kornish, 2001). In fact, when past buyers can be identified, a monopolist can price discriminate when launching a new product by either producing more of the older product, offering upgrade prices to past buyers, or buying back excess stock of the older product (Fudenberg & Tirole, 1998). We extend this literature by treating quality as an endogenous decision. This treatment of quality reflects the idea that better performing versions of a product become available for launch after a significant investment in R&D. We also
examine how the ability to develop quality affects a firm’s decision to target and “trade up” different consumer segments.\(^4\)

Other reasons why a monopolist might offer upgrades to an existing product include market growth (Ellison & Fudenberg, 2000) and the management of time inconsistency (Coase, 1972; Shankaranarayanan, 2007). In contrast to our research, this stream of literature deals with network externalities and exogenous quality.

Research in operations management considers the effect of product design on upgrade timing (Krishnan & Zhu, 2006). If the improvements that relate to the inter-operable components of an overall product are separable, firms should launch product “upgrades” frequently (Ramachandran & Krishnan, 2008). Our research does not deal with the interoperability of components; our intent is to examine products with unique functionality that can be improved through investments in product development. In a market where demand depends on durability and obsolescence, the provision of a single period “product life” as a design aspect has also been analyzed in the marketing literature (Koenigsberg, Kohli, & Montoya, 2011). We address the challenge that is faced by a firm that introduces upgraded versions of its product over time and with the goal of understanding the impact of PDC on marketing strategy (i.e., choice of target segment and pricing). This is a new question. We therefore abstract away from the aspects of product design and technology to concentrate on the effect of overall investment on product performance and a consumer’s willingness to trade up.

Information asymmetry regarding product quality has been an important area of research for durable products (Waldman, 2003). Hendel and Lizzare (1999) analyze the “lemons’ problem” (Akerlof, 1970) and show that information asymmetry can result in lower levels of trade in secondhand markets. We consider information asymmetry in the absence of secondhand markets. In a related article, Balachander and Srinivasan (1994) analyze a monopolist’s ability to signal competitive advantage to a potential entrant in a market where demand across periods is not linked. In contrast, the recipients of the signal in our model are consumers who are uncertain about product quality. In addition, the interaction of demands from one period to the next is a key aspect of the model that we consider.

Our analysis combines supply-side product development and demand-side consumer heterogeneity to understand how PDC influences a firm’s marketing strategy. In sum, the analysis demonstrates that the choice of marketing strategy is highly sensitive to the PDC of the firm, the thinking process of consumers, and the observability of quality. The following section presents the model followed by the key findings.

3. The model

We consider a monopolist firm and a heterogeneous market over two periods \(t = 1\) and \(2\). The firm sells a product of quality \(q_1\) in period 1 and \(q_2\) in period 2 where \(q_2 = q_1 + \Delta q\). The prices in each period are denoted by \(p_t\) (where \(t = 1, 2\)).

3.1. Consumer Utility

At the end of period 1, the firm can improve the quality of the product so that the second period product is better. The market consists of two types of consumers with a taste for quality \(\theta_i \forall i \in [1, H]\) with segment sizes \(1 - \lambda\) of “Highs”, type \(H\), who place a higher value on quality and \(\lambda \in [0, 1]\) of “Lows”, who place a lower value on quality \((\theta_H > \theta_L)\). A consumer’s decision in period 1 consists of choosing between a) buying now, b) waiting to buy in period 2, or c) not buying at all. As is explained below, we focus on situations where the firm serves both segments of consumers at least once. Accordingly, the first period decision boils down to buying now or waiting until period 2. When the consumer buys in period 1, she derives a surplus given by \(\theta q_1 - p_1 + \delta \theta q_1\), where \(\delta \theta q_1\) is the residual surplus in period 2, and \(\delta\) is the common discount factor. However, if she waits to buy the product in period 2, she derives a surplus of \(\delta(\theta q_2 - p_2)\). Consumer utility can then be written as \(u_1 = \max[\theta q_1 - p_1 + \delta \theta q_1, \delta(\theta q_2 - p_2)]\). When consumers are forward-looking and \(\delta > 0\), consumers are assumed to have rational expectations about the quality and price of the second period product. Conversely, by setting \(\delta = 0\) in the utility function, we represent “myopia” on the part of consumers, i.e., they think only about the present when making decisions.

To represent the consumer’s decision to buy in period 1, we define an indicator function \(I_b = 1\) if a consumer of type \(\theta_i\) buys in period 1 and zero otherwise. Similarly, the indicator function for the second period \(J_b = 1\) if a consumer of type \(\theta_i\) buys in period 2 and zero otherwise. Summarizing, we obtain the following.

\[
I_b = \begin{cases} 
1 & \text{if } u_1 = \theta q_1 - p_1 + \delta \theta q_1 - \delta(\theta q_2 - p_2) \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[
J_b = \begin{cases} 
1 & \text{if } u_2 = \theta (1 - I_b)[q_1 + \Delta q] - p_2 \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

Note that a consumer who purchased in period 1 \((I_b = 1)\) receives a residual utility of \(\theta q_1\) from the first period product in Period 2. However, she can also buy the second-generation product in Period 2 and will do so if the marginal utility is positive, i.e., if \(\theta_1[1 - I_b]q_1 + \Delta q - p_2 = \theta_1\Delta q - p_2 > 0\).

To simplify the exposition, we assume \(\theta_H = 1\) and \(\theta_L = \theta < 1\). To focus on the interesting case in which firms are torn between a) charging a high price and not serving the Lows or b) charging a low price and leaving the Highs’ premium on the table, we impose an upper bound on \(\lambda\) (the fraction of the market that is Lows’). When the Lows segment is either too large or too attractive, a supplier will mass market the product every period and treat the market as being comprised entirely of the Lows. The following bound for \(\lambda\) is derived in Appendix A.

\[
\lambda < 1 - \theta. \tag{1}
\]

In addition, the context is only interesting if the Lows are worth serving independent of the PDC of the firm. In particular, if the WTP of the Lows is too low, then the market is de facto homogeneous, composed of only the Highs. This scenario would render the question of price skimming or price penetration irrelevant. Accordingly, we restrict our attention to conditions where both the Highs and the Lows are served at least once across the two periods. To ensure that the Lows are worth serving, the second period surplus created by a quality increase \(\Delta q\) for the Highs must be less than the total surplus created by selling to the Lows for the first time. This condition is (see Appendix A):

\[
\theta_1(q_1 + \Delta q) > \Delta q \Rightarrow q_1 > \frac{1-\theta}{\theta} \Delta q. \tag{2}
\]

When this condition is violated, the firm may develop sufficient quality in period 2 such that it is worthwhile to “ignore” the Lows. In sum, we focus on the situation where profits from both types of consumers are strategically important. Eqs. (1) and (2) together are sufficient to ensure that both types of consumers are served at least once in the two periods.

\(^4\) Our focus is on categories such as consumer electronics, where the secondhand market does not have a significant effect. Yin, Ray, Gurmani and Animesh (2010) focus on markets where this is not the case.
3.1. Market demand and firm profits

The market consists of unitary aggregate demand. The demand in period 1 is given by:

\[ D_1 = \lambda \theta_b + (1 - \lambda) J_1. \]  

(3)

Because consumers are homogeneous except for their type \( \theta_i \), either all of the Lows \( (\theta_i = \theta) \) buy \( (J_0 = 1 \text{ if } \theta_1(\theta_i = \theta) \geq 0 \text{ resulting in demand } \lambda) \) or do not buy \( (J_0 = 0 \text{ resulting in no demand}). \) Similarly, either all of the Highs \( (J_1 = 1 \text{ if } \theta_1(\theta_i = 1) \geq 0 \text{ resulting in demand } 1 - \lambda) \) or do not buy \( (J_1 = 0 \text{ resulting in no demand}). \) Note that because \( \theta < 1 \), if \( J_0 = 1 \) then \( J_1 = 1 \). The demand in period 2 is given by

\[ D_2 = \lambda \theta_b + (1 - \lambda) J_1. \]  

(4)

As before, either all of the Lows buy \( (J_0 = 1 \text{ resulting in demand } \lambda) \) in period 2 or do not buy \( (J_0 = 0 \text{ resulting in no demand}). \) Similarly, either all of the Highs buy \( (J_1 = 1 \text{ resulting in demand } 1 - \lambda) \) or do not buy \( (J_1 = 0 \text{ resulting in no demand}). \)

The consumers and the firm are risk-neutral and have a common discount factor \( \delta \) (except in the case of myopic consumers). We assume that there is no after-market for previous generation products and that they are disposed of at zero cost when a new generation is purchased. The firm invests in product development at the beginning of period 2. Accordingly, total profits are given by

\[ \pi = (p_1 - c_1) D_1 + \delta (p_2 - c_2) D_2 - C(\Delta q) \text{ where } C(\Delta q) = \frac{\Delta q^2 + 2.q_1.\Delta q}{2.\alpha}. \]

In this case, \( C(\Delta q) \) is the cost of developing an improvement in quality of \( \Delta q \) and \( \alpha \) reflects the firm’s PDC \((\alpha > 0)\). The parameter \( \beta \) is the technological cost factor. This factor is positive when the cost of product development increases as the existing quality \( q_1 \) approaches a technological frontier. To simplify, we assume that the firm’s level of investment in product development becomes common knowledge at the end of period 1. This assumption is reasonable when consumers learn about the firm’s product development capability from their interaction in period 1. Rational expectations imply that consumers correctly infer the profit maximizing quality increase.

The marginal cost \( c_2 = c_1 + \rho \Delta q \) (where \( \rho \equiv [0, \theta] \)) increases in proportion to the quality that is developed \( (\Delta q) \) and \( c_1 = c > 0 \).

Furthermore, without loss of generality, we assume \( c = 0.6 \).

Consequently, the firm’s profit across two periods is written as:

\[ \pi = p_1 D_1 + \delta \left[ (p_2 - \rho \Delta q) D_2 - \frac{\Delta q^2 + 2q_1.\Delta q}{\delta.\alpha} \right]. \]  

(6)

The parameters \( \rho \) (the factor for marginal cost increase in period 2) and \( \delta \) (the discount factor) are assumed to be common knowledge. Note that this framework can be extended to the case when the outcome of product development or R&D is uncertain.\(^7\)

4. Consumers are myopic

We first consider “myopic” consumers who maximize utility in each period and do not consider the future. Myopia may be a reasonable representation of consumer behavior when either the firm’s PDC or its plans are unknown. On the one hand, PS marked Sony’s entry into the market for video game consoles. At the time of the launch, consumers focused on the immediate incremental benefits of the PS. In situations such as this, buyers are less likely to assess the benefits of a “theoretical” next generation when buying the first product. On the other hand, consumers develop rational expectations about the future generations of products when firms have a “history” of introducing improved versions over time. For example, consumers are likely to develop expectations about the future generations of iPods and iPhones over time based on the historical frequency of upgrades that have been introduced by Apple. We analyze the latter situation by considering forward-looking consumers in Section 5.

4.1. Quality is observable

We first examine the case when consumers can observe firm type. This case is a helpful benchmark to understand how asymmetric information affects the market. When quality is observable, myopic consumers buy whenever they obtain a positive net surplus by buying and using the new product.

The first period utility of the two types of myopic consumers is given by the utility functions

\[ u_i(\theta_i = \theta) = \theta q_1 - p_i \text{ for the Lows, } u_i(\theta_i = 1) = q_1 - p_i \text{ for the Highs}. \]

Similarly, utility of buying the product for the first time in period 2 for the Lows and the Highs are respectively given by \( \theta (q_1 + \Delta q) - p_2 \) and \( q_1 + \Delta q - p_2 \). However, because a consumer who purchased in period 1 receives a residual utility of \( \theta q_1 \) (or \( q_1 \)) from the product she owns, the marginal utility of purchasing in period 2 is \( \Delta q - p_2 \). Therefore, with the indicator functions \( J_0 \) and \( J_1 \), the second period utilities of the Lows and the Highs are written as:

\[ u_2(\theta_i = \theta) = \theta (1 - J_0) q_1 + \Delta q - p_2 \text{ for the Lows, } u_2(\theta_i = 1) = (1 - J_1) q_1 + \Delta q - p_2 \text{ for the Highs}. \]

Similarly, we use the indicator function \( f \) to capture the consumer’s decision to buy in period 2:

\[ J_0 = \{ 1 \text{ if } \theta q_1 \Delta q - p_2 \geq 0 \text{ and } q_1 \Delta q - p_2 \geq 0 \text{ otherwise } \}. \]

Similarly, we use the indicator function \( f \) to capture the consumer’s decision to buy in period 2:

\[ J_1 = \{ 1 \text{ if } q_1 + \Delta q - p_2 \geq 0 \text{ otherwise } \}. \]

4.1.1. The extensive form of the game

The game we analyze has two stages:

Stage 1 The firm chooses the price \( p_1 \) for the first period product, and consumers evaluate the offer that was made by the firm based on the quality \( q_1 \).

Stage 2 At the beginning of period 2 (\( t = 2 \)), the firm invests in product development to deliver a quality improvement \( \Delta q \). Both the investment and the developed quality are observable to consumers. The firm offers a price \( p_2 \) for the second period product, following which consumers decide whether to buy.

As noted earlier, we consider a market where both types of consumers are served at least once (\( i_0 + i_1 > 0 \) and \( j_0 + j_1 > 0 \)) and the firm sells in both periods (\( i_0 + j_0 > 0 \) and \( i_1 + j_1 > 0 \)). The firm maximizes the objective function \( \max_{p_1, p_2} \pi \) where \( \pi \) is
given by Eq. (6). There are four possible market outcomes where both types of consumers are served at least once and the firm sells in both periods. Each outcome is associated with the specific values of $D_1$ and $D_2$ and a set of constraints, which are as follows.

a) Market penetration: Both segments buy in period 1 ($I_0 = 1, I_1 = 1$) but only the Highs buy in period 2 ($J_0 = 0, J_1 = 1$) under the following conditions (or “pricing constraints”) $p_1 \leq \theta q_1$ and $\theta \Delta q < p_2 \leq \Delta q$. The firm decision problem is therefore given by

$$\pi_{MP} = \max \pi \text{ s.t. } \theta q_1 - p_1 \geq 0, \text{ and } \Delta q - p_2 \geq 0.$$  \hspace{1cm} (9)

b) High-end focus and then mass market: The Highs buy in period 1 ($I_0 = 0, I_1 = 1$) but both segments buy in period 2 ($J_0 = 1, J_1 = 1$) under the conditions $\theta q_1 < p_1 \leq q_1$ and $p_2 \leq \Delta q < \theta (q_1 + \Delta q)$. The firm decision problem is therefore given by

$$\pi_{HFM} = \max \pi \text{ s.t. } \theta q_1 - p_1 \geq 0, \text{ and } \Delta q - p_2 \geq 0.$$  \hspace{1cm} (10)

c) Market inversion: The Highs buy in period 1 ($I_0 = 0, I_1 = 1$) and the Lows buy in period 2 ($J_0 = 1, J_1 = 0$) under the conditions $\theta q_1 < p_1 \leq q_1$ and $\Delta q - p_2 \leq \theta (q_1 + \Delta q)$. The firm decision problem is therefore given by

$$\pi_{MI} = \max \pi \text{ s.t. } \theta q_1 - p_1 \geq 0, \text{ and } \Delta q - p_2 \geq 0.$$  \hspace{1cm} (11)

d) Mass market: Both segments buy in both periods ($I_0 = 1, I_1 = 1, J_0 = 1, J_1 = 1$). The conditions are $p_1 \leq \theta q_1$ and $p_2 \leq \Delta q$. The firm decision problem is therefore given by

$$\pi_{MM} = \max \pi \text{ s.t. } \theta q_1 - p_1 \geq 0, \text{ and } \theta \Delta q - p_2 \geq 0.$$  \hspace{1cm} (12)

Table 1 summarizes the above.

We now solve the firm’s constrained optimization problem for each of the decision alternatives and then compare the profit values across the alternatives. Note that the optimal marketing strategies depend on the firm’s level of PDC. Proposition 1 describes the optimal strategies of the firm as a function of $\alpha$, the firm’s PDC. Please refer to the appendix for all the proofs.

**Proposition 1.** Myopic consumers: Under complete information, a firm having PDC higher than a threshold ($\alpha \geq \alpha_{M}$ where $\alpha_{M} = \sqrt{2q_{1} \theta p_{1}} \frac{\theta}{1 - \theta p_{1}}$) serves the Highs in both periods and the Lows in period 2 (“High-end focus and then mass market” strategy). A firm having a PDC that is lower than the threshold ($\alpha < \alpha_{M}$) serves only the Highs in period 1 and only the Lows in period 2 (“Market inversion strategy”).

Proposition 1 demonstrates that the PDC of the firm has a significant influence on its optimal marketing strategy: the price and the segment(s) it targets. A firm with a high PDC ($\alpha \geq \alpha_{M}$) can develop higher quality at a lower cost than a firm with a low PDC ($\alpha < \alpha_{M}$).

In fact, the firm can profitably develop the next generation with a price-quality offer such that the Highs trade up. In period 2, therefore, this type of firm sells the new generation product to both the Highs and the Lows. We call this a “high-end focus and then mass market” strategy. Note that both the Highs and the Lows use state-of-the-art technology in period 2. A firm with high PDC develops a quality and charges a price that optimizes profits from second period sales to both the Highs and the Lows, the reservation price being equal to the WTP of the Highs. Nevertheless, the optimal price is sufficiently low such that the Lows buy for the first time in period 2. This sale to the Lows is an important source of revenue for the firm in period 2 when $\alpha \geq \alpha_{M}$. Interestingly, the Lows entering the market in period 2 realize a positive surplus because they have a higher WTP for the second generation than the Highs, who already own a first-generation product.

A low PDC firm, however, cannot create significant value from the second-generation product due to the high cost it incurs to increase product quality. Consequently, the primary focus of a low PDC firm is to maximize its earnings from the endowed quality of the first-generation product. The firm accomplishes this by intertemporally discriminating between the Highs and the Lows, i.e., it charges a high price and sells to the Highs in period 1 and then reduces the price to realize the demand from the Lows in period 2. We refer to this strategy as “market inversion”; in period 2, the Highs continue to use the first-generation product, while the consumers who obtain lower value from the improved version of the product (the Lows) purchase and use the second-generation product. Market inversion is optimal for a low PDC firm because the incremental value of the second period product to the Highs is less than its “total value” for the Lows.

In Fig. 1, the threshold $\alpha_{M}$ divides the parameter space as a function of the optimal strategy for the supplier. Firms with a high PDC spend more on product development in period 2 and develop higher quality products. Furthermore, the higher the quality that is developed, the opportunity cost to serve the Lows increases (the price reduction that is needed to sell to the Lows is higher). Conversely, if either $\theta$ or $\lambda$ increases, the firm has more incentive to serve the Lows. This explains why in the upper left area of the Figure (Zone 1), the Lows are of high importance (they are the only consumers who are served in period 2). In contrast, in the upper right zone (Zone 2), the primary sources of revenue are the Highs. However, the Lows enter the market in period 2. They find the second period
product attractive at the price which extracts all of the surplus from the Highs.

To summarize, we find that firms with a low PDC tend to market more broadly after the product launch. In contrast, the pricing of firms with a high PDC is driven by the WTP of the Highs.

4.2. Quality is unobservable

We now consider a situation where the firm cannot convey credible information about product quality before consumers make their first purchase. Similar to the previous section, consumers do not think about second period benefits when making their first period decision. They may, however, infer information about product quality from the firm’s actions such as its offer of price. Consequently, a firm has the opportunity to ‘signal’ product quality through its price.\(^\text{12}\)

The modeling setup we use to investigate the case of unobservable quality is as follows.

4.2.1. Firm types

We consider two types of firms based on their PDC. A firm with a high PDC incurs a lower cost to improve quality, i.e., \(\alpha = \alpha_H\), while the low PDC firm incurs a higher cost to improve quality, i.e., \(\alpha = \alpha_L\), where \(\alpha_H > \alpha_L\).\(^\text{13}\)

The firm with a high PDC has a product in period 1 of quality \(q_H\). The firm with a low PDC has a product of lower quality \(q_L\) in period 1 (i.e., \(q_H > q_L\)). Furthermore, we assume that \(\alpha_H > \alpha_L(q_1 = q_H)\) and \(\alpha_M(q_1 = q_L) > \alpha_L\), which ensures that were quality observable, the optimal strategy of the high PDC firm would be High-end focus and then mass market while that of the low PDC firm would be Market inversion.\(^\text{14}\) Note that when either \(\alpha_H > \alpha_L > \alpha_M(q_1 = q_H)\) or \(\alpha_M(q_1 = q_L) > \alpha_H > \alpha_L\), both types of firms pursue exactly the same strategy under complete information. In those cases, the high PDC firm is unable to signal its quality. We are interested in parameter conditions where the strategies of the two types of firms, under complete information, are different.

4.2.2. The extensive form of the game

Incorporating the asymmetric information about firm type, we have a two-stage game under incomplete information, which proceeds as follows (see Fig. 2):

Stage 1 Nature chooses the firm type, either high (\(\alpha_H\)) or low (\(\alpha_L\)) quality. As noted earlier, firm type is perfectly correlated with PDC. The firm then chooses \(p_1\) for the first period product. The consumers cannot evaluate quality (nor do they observe the firm type). Consequently, consumers need to form beliefs about the quality of the product (and the firm’s PDC) based on the offer made that is by the firm. We represent these beliefs with \(\mu\), the probability that consumers believe that the firm has a high PDC. After buying and using the product in period 1, consumers learn about its quality. Consequently, consumers know the firm’s type at the start of period 2.

Stage 2 At the beginning of period 2 (\(t = 2\)), the firm invests in product development to deliver additional quality \(\Delta q\). Because the firm type is known at the beginning of period 2, consumers know the quality of the product prior to purchase in period 2. The firm offers a price \(p_2\) for the second period (next generation) product, and the consumers decide whether to buy it.

The rest of the assumptions are identical to those that were made for the case of complete information that was discussed in Section 4.1.

---

\(^{12}\) Many researchers have considered various means of signaling quality: advertising (Milgrom & Roberts, 1986), price (Choi, 1998) and warranties (Soberman, 2003). Price as a signal may be distorted either upward (e.g., Choi, 1998) or downward (Milgrom & Roberts, 1986) to signal higher quality.

\(^{13}\) This is a reasonable assumption because the quality of the original product is also a result of the efforts of the developers, technical personnel and facilities that were employed by the firm for the development of the second-generation product. We assume that the quality of those resources is highly correlated over time.

\(^{14}\) Note that because \(q_H > q_L\), we have \(\alpha_L < \frac{\Delta q}{p_2} \left( \frac{\rho - \lambda \theta}{\rho - \lambda \theta - \beta} \right) \leq \alpha_H\).
The objective is to identify a ‘separating’ equilibrium in which the firm maximizes its profit and the consumers receive the quality that they believe they are purchasing (a situation in which a consumer believes she is purchasing a high quality product but actually receives low quality does not constitute an equilibrium). In other words, the actions of a firm with a high PDC that sells a high quality product are constrained by consumers’ inferences about quality. We introduce this as a constraint in the high PDC firm’s optimization problem. This constraint leads to a standard signaling game in which the uninformed player (the consumer) makes an inference about the type of the informed player (the firm) that is based on the latter’s action.

A key assumption in the analysis is that consumers know that the firm can have two levels of PDC and the absolute levels associated with each type are common knowledge, i.e., consumers know the firm’s cost function for increasing quality, and they also know that the $\alpha = \alpha_H$ or $\alpha_L$. To signal higher quality, the high PDC firm changes its first period offer as explained in Proposition 2.

**Proposition 2.** When product quality is unobservable, the high PDC firm signals its type by offering a lower price ($p^H_1 = q_H$) to myopic consumers in period 1 compared to the case of complete information. The actions of a firm with a low PDC are unaffected.

Proposition 2 shows that a firm with a high PDC charges a lower price compared to the price when product quality is observable. In fact, Proposition 2 implies that this price is the same as the complete information price of a firm with a low PDC, which results in a “tie.” However, the high PDC firm can drop the price by an arbitrarily small amount $\varepsilon > 0$ (i.e., $p^H_1 = q_H - \varepsilon$), at which the low PDC firm is strictly worse off, thereby breaking the tie. The low PDC firm does not mimic this strategy because it is unprofitable; thus, its actions identify it as a firm with a low PDC because it charges the price (and employs the strategy) that is used under complete information.

Signaling allows a high PDC firm to identify itself in period 1. In period 2, it charges the complete information price because its type is known. This implies that a high PDC firm will charge lower prices for new (first-generation) products when quality is unobservable. The impact of a high PDC firm charging lower prices is summarized in Proposition 3.

**Proposition 3.** When quality is unobservable and the quality in period 1 is above a threshold ($q_H \geq q_L / \theta$), the high PDC firm covers the market early to signal its type and restricts the sales of the new generation product to the Highs in period 2 (‘Market penetration’ strategy). Note that when quality is observable, a firm with a high PDC does not employ first period “Market penetration.”

Under complete information, a firm with a high PDC focuses on the Highs in period 1 and in period 2, it sells the product broadly (to the Highs and the Lows). Conversely, when consumers cannot observe quality, the strategy of the high PDC firm changes. Proposition 3 implies that the high PDC firm signals its type by offering a lower price in period 1 and this makes the first-generation product attractive to the Lows as well as the Highs. Then in period 2, the high PDC firm restricts sales to the Highs.

In summary, asymmetric information (and the, corresponding need to signal) causes a firm with high PDC to completely reverse the launch strategy it would use were quality observable (broad then narrow versus narrow then broad).

It is also useful to underscore why a firm with a low PDC does not mimic the pricing of the firm with a high PDC. Specifically, there is nothing to stop a low PDC firm from copying the high PDC firm’s price: this would lead to increased first period demand as the firm would sell to the Lows as well as to the Highs. However, the main objective of a low PDC firm is to maximize earnings from the *endowed quality* of the first-generation product (high product development costs make it infeasible to offer a significant quality improvement in period 2). Were the low PDC firm to mimic the offer of the high PDC firm, it would lose the ability to intertemporally discriminate between the Lows and the Highs. Moreover, a small improvement in product quality is naturally offered with a very low price to consumers who already have a product (all consumers have a first-generation product with the high PDC strategy described in Proposition 3).

The model highlights who gains and loses when the quality of products is unobservable. The primary loser is the firm with a high PDC. In other words, signaling costs money, and firms with a high PDC (and

---

$^{15}$ When $\alpha_H = \alpha_L$, the firm has low PDC and the optimal strategy is market inversion.
hence higher quality first period products) need to signal. The primary beneficiaries of asymmetric information are the Higns. When product quality is high, the Higns are able to buy products with an attractive price that is essentially discounted by the seller.

5. Consumers are forward-looking

In this section, we evaluate firm decisions in the same market, but we do so in cases in which the consumers are “forward-looking”, i.e., the consumers maximize their utility across the two periods based on rational expectations about the firm’s price and quality decisions in period 2. As in the case of myopic consumers, we first consider the case where quality is observable.

5.1. Quality is observable

A forward-looking consumer with rational expectations buys the product in period 1 if and only if she expects a higher net surplus as compared to waiting and purchasing in period 2. For the Lows and the Higns respectively, this behavior implies \( \theta_1 - p_1 + \delta \theta_1 \geq \delta (\theta_2 - p_2) \) and \( q_1 - p_1 + \delta q_1 \geq \delta (q_2 - p_2) \). At this point, we set forth the utility for the Lows and the Higns in terms of the product improvement \( \Delta q \) that is made to the second-generation product:

\[
U_i(\theta) = \max(\theta_1 - p_1 + \delta \theta_1, \delta (\theta_1 + \Delta q - p_2))
\]

Using the conventions of Section 4.1, we write the corresponding indicator functions in the case of forward-looking Lows (Higns) as

\[
I_o = \begin{cases} 1 & \text{if } \theta_1 - p_1 \geq \delta (\theta_1 + \Delta q - p_2) \text{ and } q_1 - p_1 \geq \delta (q_1 + \Delta q - p_2) \text{ for the Lows;} \\ 0 & \text{otherwise} \end{cases}
\]

\[
l_i(\theta) = \begin{cases} 1 & \text{if } q_1 - p_1 \geq \delta (q_1 + \Delta q - p_2) \text{ for the Higns;} \\ 0 & \text{otherwise} \end{cases}
\]

The second period utility from purchase is the same as it is for myopic consumers that is given by Eq. (8). Aside from the value that consumers obtain by using the first period product, the model is very similar to the specification of Section 4.

The market demand and firm profit functions are identical to those of Section 4.1.

5.1.1. The extensive form of the game

The timing of the game is identical to that of Section 4.1. Only the basis for consumer decisions changes when consumers think ahead.

While the firm strategies remain the same as in Section 4.1 and Eq. (6), the necessary conditions (or pricing constraints) for each strategy are different, this difference reflects the forward-looking decision basis that consumers use to make decisions.

a) Market penetration: \((I_o = 1, l_i = 1), \text{ and } (J_o = 0, J_1 = 1)\). The necessary conditions are \( \theta_1 - p_1 \geq \delta (\theta_1 + \Delta q - p_2) \Rightarrow p_1 \leq \theta_1 - \delta (\theta_1 + \Delta q - p_2) \) and \( \theta_1 < \theta_1 + \delta (\theta_1 + \Delta q - p_2) \). The firm decision problem, therefore, is

\[
\pi_{mb} = \max_{p_1, p_2, \Delta q} \theta_1 - \delta (\theta_1 + \Delta q - p_2) - p_1 \geq 0 \text{ and } \Delta q - p_2 \geq 0.
\]

b) High-end focus and then mass market: \((I_o = 0, l_i = 1), \text{ and } (J_o = 1, J_1 = 1)\). Here, the conditions are \( \theta_1 - \delta (\theta_1 + \Delta q - p_2) \leq p_1 \leq \theta_1 - \delta (\theta_1 + \Delta q - p_2) \) and \( p_2 \leq \Delta q < \theta_1 + \Delta q \). The firm decision problem, therefore, is

\[
\pi_{bf} = \max_{p_1, p_2, \Delta q} \theta_1 - \delta (\theta_1 + \Delta q - p_2) - p_1 \geq 0 \text{ and } \Delta q - p_2 \geq 0.
\]

c) Market inversion: \((I_o = 0, l_i = 1), \text{ and } (J_o = 1, J_1 = 0)\). The necessary conditions are \( \theta_1 - \delta (\theta_1 + \Delta q - p_2) \leq p_1 \leq \theta_1 - \delta (\theta_1 + \Delta q - p_2) \) and \( \Delta q < p_2 \leq \theta_1 + \Delta q \). The firm decision problem, therefore, is

\[
\pi_{bi} = \max_{p_1, p_2, \Delta q} \theta_1 - \delta (\theta_1 + \Delta q - p_2) - p_1 \geq 0 \text{ and } \Delta q - p_2 \geq 0.
\]

d) Mass market: \((I_o = 1, l_i = 1), \text{ and } (J_o = 1, J_1 = 1)\). Here, the conditions are \( p_1 \leq \theta_1 - \delta (\theta_1 + \Delta q - p_2) \) and \( p_2 \leq \Delta q. \) The firm decision problem, therefore, is

\[
\pi_{mm} = \max_{p_1, p_2, \Delta q} \theta_1 - \delta (\theta_1 + \Delta q - p_2) - p_1 \geq 0 \text{ and } \Delta q - p_2 \geq 0.
\]

As in the case of myopic consumers, we solve the firm decision problem for each of the above strategies and compare the outcomes to find the strategy that maximizes profit. Proposition 4 describes the optimal strategies for the firm as a function of \( \alpha \) when consumers are forward-looking.

Proposition 4. Forward-looking consumers: Under complete information, a firm having a PDC higher than the threshold \( \alpha_0 > \alpha_b \) (where \( \alpha_b = \frac{2q_1}{\lambda_1 + (\lambda_2 - \lambda_1)\beta} \)) sells to both the Higns and the Lows in period 1 and only to the Higns in period 2 ("Market penetration" strategy). A firm having a PDC of less than the threshold (\( \alpha < \alpha_b \)) serves the Higns in both periods and the Lows only in period 2 ("High-end focus and then mass market" strategy).

Proposition 4 shows that forward-looking behavior by consumers leads to a lower WTP for the first period product because consumers weigh the benefit of consuming today against the cost of waiting to buy in period 2. The impact of this behavior by consumers hurts the firm: forward-looking behavior creates a durable goods monopoly problem (Coase, 1972). The Higns know that they would trade up in period 2, implying that the first period product is purchased based only on its consumption in period 1. Because the firm is forced to charge a lower price in period 1 (due to the forward-looking behavior of consumers), both types of consumers buy the first-generation product. Under this scenario, the optimal strategy for the high PDC firm is to develop a higher quality second-generation product that is targeted at the Higns alone and to charge a correspondingly higher price for it.

Conversely, the strategy shifts from Market inversion to High-end focus and then mass market when a firm has a low PDC. As with a high PDC firm (when consumers are forward-looking), the price that consumers are willing to pay in period 1 is driven downwards. Independent of the strategy that is employed by the firm, this drop in price hurts firm profits. However, when a firm has a low PDC, the incremental increase in product quality in period 2 is small (because of the high cost of product development). Accordingly, a firm with a low PDC has an incentive to get as much money as it can from its “endowed” first period quality. It accomplishes this by selling only to the Higns in period 1. In a sense, this behavior allows the low PDC firm to charge higher prices in both periods. In period 2, it proves beneficial for the firm to sell the new generation broadly. The Higns are interested in the improved performance of the new generation product and the WTP of the Lows is driven by the total benefit that is provided by the product as they did not buy in period 1.

In sum, the most important effect that is created by consumers’ forward-looking behavior is that consumers (the Higns and the Lows) have a lower WTP in period 1. This phenomenon has two consequences. First, it reduces first period prices and constrains the firm,
independent of PDC. Second, forward-looking behavior by consumers brings the Lows into the market sooner when a firm has a high PDC. The high PDC firm wrestles with the choice of charging a price that the Lows find acceptable and extracting surplus from the Highs. By serving the Lows in period 1, the high PDC firm unshackles itself in period 2 and is able to charge the Highs a high maximum price for a “significantly improved” second-generation product.

In contrast, the firm with a low PDC extracts the maximum price in period 1 (from the Highs) and then serves both the Highs and the Lows in period 2. Interestingly, when consumers are forward-looking, the Highs buy in every period independent of the PDC of the firm. In contrast, when consumers are myopic, the Highs remain with the first-generation product when the firm has a low PDC.

5.2. Quality is unobservable

As in Section 4.2, we consider a situation where the firm cannot convey credible information about product quality before consumers make their first purchase. Similar to the previous section, consumers think about their expected second period benefits while making their period 1 decision. Because consumers will make inferences about product quality when it is not observable, a firm may be able to signal its quality through pricing.

5.2.1. Firm types

Aside from the following assumption, firm types are identical to those for myopic consumers as described in Section 4.2. We assume that $\alpha_H > \alpha_F(q_1 = q_H)$ and $\alpha_F(q_1 = q_L) > \alpha_L$ to ensure that the optimal strategies of the high PDC firm and low PDC firms are Market penetration and High-end focus and then mass market, respectively, when quality is observable. When either $\alpha_H > \alpha_L > \alpha_L(q_1 = q_H)$ or $\alpha_F(q_1 = q_L) > \alpha_H > \alpha_L$, both types of firms pursue the same strategy under complete information, and the high PDC firm cannot signal its quality. Similar to Section 4.2, parameter conditions where the strategies of the two types of firms under complete information are different lead to signaling.

5.2.2. The extensive form of the game

The timing of the game is identical to that of Section 5.1 except for the following: in Stage 1, Nature chooses the firm type either high ($\alpha_H$) or low ($\alpha_L$) quality and consumers form beliefs with probability $\mu$ that the firm has a high PDC. Similar to Section 4.2, consumers become informed about the firm’s type before the second period. Proposition 5 explains the optimal firm strategy when consumers are forward-looking but cannot observe product quality.

**Proposition 5.** When quality is unobservable and consumers are forward-looking, the high PDC firm signals its quality by offering a lower price ($p_{1H}^f < p_{1L}^f$) than under complete information, but using the same “Market penetration” strategy as under complete information. The actions of a firm with a low PDC are unaffected.

When quality is unobservable, a high PDC firm offers a lower price in period 1 to signal its quality. This finding echoes the equilibrium that obtains when consumers are myopic and quality is unobservable (see Proposition 2). There is an important difference, however. Other than charging a lower price (due to the need to make the signal informative), the launch strategy of the high PDC firm is unaffected by the observability of quality.\(^{18}\) That is, a firm with a high PDC utilizes a strategy of Market penetration independent of whether quality is observable. This phenomenon indicates an important interaction between the way that consumers think about their purchase decision and the observability of quality. When consumers think myopically about decisions, unobservable quality changes the optimal strategy of a firm with a high PDC (the high quality firm has $q_H > q_L / \theta$) from High-end focus and then mass market to Market penetration. In contrast, when consumers are forward-looking, the strategy of a firm with a high PDC is unaffected; only the price is different. Perhaps the most interesting finding with regard to the impact of consumers thinking ahead (for their first purchase decision) is the effect it has on the Lows when firms have a high PDC. Forward thinking behavior brings the Lows into the market sooner when a firm has a high PDC. The basic force driving this finding is the desire of the high PDC firm to capitalize on selling the second-generation product to the Highs. The sooner the Lows are served (and effectively removed from the market), the greater the flexibility of the high PDC firm to capitalize on the high WTP of the Highs for the next sale.

6. Conclusions

The objective of this paper is to examine how the PDC of a firm that sells upgradable durable goods affects marketing and product development decisions in a market where consumers have heterogeneous valuations for performance. Under complete information, the optimal marketing strategy of a firm with a high PDC revolves around selling (and re-selling) to high WTP consumers. In such cases, extracting value for new product generations from high WTP consumers is critical. This finding may explain why well-known makers of high-end consumer electronics (such as Sony, Apple, and Royal Philips Electronics) and sporting equipment such as Völkl (skis) and Graf (hockey skates) make ongoing investments in product development to develop new generations on a regular basis. The firm’s efficiency at developing higher levels of performance makes offering new generations on a regular basis attractive.

In contrast, the optimal strategy of firms with a low PDC is to focus on expanding the market after serving high WTP consumers in period 1. Because firms with a low PDC improve their products less over time, they expand the market over time to optimize profitability. Under certain conditions, the strategy of selling to as many consumers as possible leads to “market inversion”: the firm serves high WTP consumers with early versions of the product, and low WTP consumers buy improved versions. In this case, high WTP consumers continue to use old technology, even when a newer version of the product is available.

Typically, a firm sells both high and low quality products, respectively, to high and low WTP consumers (see, for example, Maskin & Riley, 1984). A phenomenon that is similar to market inversion known as “vintage effects” has been noted in the literature (see, for example, Bohlin, Golder, & Mitra, 2002). When two competing firms enter a market sequentially, the follower uses more recent technology than an incumbent who has committed to older technology. Thus, an established market leader is sometimes observed to use less efficient technology (an earlier vintage) than a new entrant. Of course, consistent with our findings, this situation occurs when the new generation product offers a small advantage over the original technology.

In some cases, it is difficult for consumers to evaluate the quality of a new product. For example, many buyers of consumer electronics or specialized sporting equipment appreciate performance only after having used the product. This scenario occurs when an unknown firm or a firm without a track record launches a product that provides unique value. In these situations, a firm with a high PDC will signal its higher performance by offering its products at a lower price than the optimal price that is charged by a firm with a low PDC. This phenomenon leads to mass sell-in for the new product in period 1. In the second period, the firm restricts the sales of the new generation product to high WTP consumers. Basically, this behavior is a reversal of the strategy that is used by a firm with a high PDC under complete information.
It is important to note that the offering of a firm with a low PDC is not affected by the unobservability of product quality. Why does a low PDC firm not mimic the strategy of a high PDC firm and pretend to be higher quality? The reason is that a firm with a low PDC optimizes its profit by intertemporally discriminating between high and low WTP. In other words, the low PDC firm needs to focus on maximizing earnings from the \textit{endowed quality} of the first-generation product because the second-generation product is only marginally better than the first-generation product. If the low PDC firm mimics the high PDC firm, it loses the ability to intertemporally discriminate between low and high WTP consumers. The low PDC firm is so constrained because its type is common knowledge by the time consumers decide whether to buy in the second period.

We also examine how forward-looking behavior on the part of consumers affects the launch strategy of firms as a function of a firm’s PDC. We show that forward-looking behavior causes a firm with a high PDC to implement a marketing strategy that better aligns the consumer valuations of product quality and the relative quality of the products that they consume. In other words, the consumers’ rational expectations about future products cause a high PDC firm to sell earlier versions of products more broadly (across consumer types) and improved versions only to high WTP consumers. In period 2, low WTP consumers do not trade up, while high WTP consumers trade up to the second-generation product. Finally, we examine how the strategy of a high PDC firm is affected by “hidden quality” when consumers think ahead. Interestingly, “hidden quality” does not qualitatively alter a high PDC firm’s marketing strategy; its only effect is to reduce the price of high quality products in period 1, as a high PDC firm “signals” its quality to the market.

Our research shows that product development capability has a significant impact on the marketing strategy that a firm should adopt to launch a product in a market where there is significant heterogeneity with regard to consumers’ valuation of product performance. In general, firms develop their marketing strategies after products have been developed. After all, why spend time developing a marketing strategy for a product that is not ready?

Our analysis shows that this approach to product development and marketing is not optimal. In particular, the optimal price of the first-generation product is affected significantly by a) the PDC of the firm, b) whether consumers assess future benefits when making current decisions, and c) the ease with which consumers can evaluate the quality of a new product. Moreover, the optimal level of investment in product improvement is impacted by the optimal pricing and expected demand for the improved product. When a firm develops sequential generations of a durable product for a heterogeneous market, our study underscores the benefits of utilizing an integrated approach for these two activities. Table 2 summarizes the optimal strategies of the firm under the different conditions that are analyzed in our model.

Our study also suggests new areas of future research. One of those areas is to analyze how competition between firms might affect the results. The model we have presented concerns a monopolistic seller that sells to two segments over time that are different in terms of willingness to pay. Using this structure as a starting point, competition might enter into the problem in two ways. First, competition might result in a reduction of the premium that consumers are willing to pay for a firm’s product. This reduction in premium would reduce firm incentives to invest in product development. Second, firms might choose to focus on high or low WTP consumers as a function of their product development capability. Another interesting avenue to explore would be the reactions of an incumbent that has the option of selling in both periods but is faced with a second period entrant whose quality might be high or low.

### Appendix A

#### Derivation of condition 1

Consider the firm decision in period 1. If the firm sells only to the Highs \((l_0 = 0, l_1 = 1)\), the firm chooses a price \(p_1 = q_1\) and obtains a demand of \(D_1 = 1 - \lambda\). On the other hand, if it sells to both types of consumers \((l_0 = l_1 = 1)\), it chooses a price \(p_1 = \theta q_1\) and obtains a demand of \(D_1 = 1\). We can see that the firm earns less profit from selling only to the Highs unless \((1 - \lambda)(q_1 - c) \geq \theta q_1 - c\), where \(c\) is the marginal cost. This condition is trivially satisfied as long as \(1 - \lambda \geq \theta\).

#### Derivation of condition 2

Due to condition 1, the firm serves only the Highs in period 1 \((l_0 = 0, l_1 = 1)\). Now consider the firm decision in period 2. If \(\Delta q \geq \theta (q_1 + \Delta q)\) in violation of condition 2, the firm can possibly earn higher profits by charging a higher price that is equal to \(\Delta q\) at which only the Highs buy \((l_0 = 0, l_1 = 1)\), leading to a demand \(D_2 = 1 - \lambda\) because \(1 - \lambda \geq \theta\) (condition 1). The Lows are therefore not served.

However, when \(\Delta q < \theta (q_1 + \Delta q)\), if the firm charges a higher price equal to \(\theta (q_1 + \Delta q)\) only the Lows buy \((l_0 = 0, l_1 = 1)\), leading to a demand \(D_2 = \lambda\). Alternatively, it can charge a lower price that is equal to \(\Delta q\) at which both types of consumers buy \((l_0 = 1, l_1 = 1)\), leading to a higher demand \(D_2 = 1\). Therefore, if the firm has any demand at all in period 2, it is guaranteed that the Lows buy when \(\Delta q < \theta (q_1 + \Delta q)\). In other words, if \(D_2 > 0, l_0 = 0\) is assured under the sufficient (but not necessary) conditions 2 and 1.

#### Proof of Proposition 1

We solve the firm’s constrained optimization problem for each of the decision alternatives and then compare the profit values across all of the strategic alternatives. Fig. 3 illustrates a characterization of the market outcomes as a function of the prices of the two periods. We apply the Kuhn–Tucker necessary and sufficient conditions for global maximum for each of the four strategies below, as the objective function is concave and the constraints are linear.

#### Market penetration

Substituting \((l_0 = l_1 = 1) \rightarrow D_1 = 1\) and \((l_1 = 1, l_0 = 0) \rightarrow D_2 = 1 - \lambda\), in Eqs. (6) and (9), we obtain the Lagrangian

\[
L = p_1 + \delta(p_2 - p\Delta q)(1 - \lambda) - \delta \frac{\Delta q^2 + 2\theta q_1 \Delta q}{2\alpha} - \mu_1(p_1 - \theta q_1) - \mu_2(p_2 - \Delta q)
\]

where \(\mu_1 \geq 0\) and \(\mu_2 \geq 0\).

The Kuhn–Tucker conditions are

#### Table 2

<table>
<thead>
<tr>
<th>Consumers (l_0 \rightarrow l_1)</th>
<th>Myopic</th>
<th>Forward-looking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality (l_1 \rightarrow l_2)</td>
<td>Firm 1</td>
<td>Firm 2</td>
</tr>
<tr>
<td>Observable</td>
<td>Hidden</td>
<td>Observable</td>
</tr>
<tr>
<td>High PDC</td>
<td>High-end focus and then mass market</td>
<td>Market penetration</td>
</tr>
<tr>
<td>Low PDC</td>
<td>Market inversion</td>
<td>Market inversion</td>
</tr>
</tbody>
</table>

#### Marginal condition

\[
\frac{\partial L}{\partial p_1} - \delta (1 - \lambda) - \frac{\partial q_1 \Delta q}{\alpha} \leq 0
\]

#### Complementary slackness

\[
\frac{\partial L}{\partial \mu_1} = p_1 - \theta q_1 = 0
\]

\[
\frac{\partial L}{\partial \mu_2} = p_2 - \Delta q = 0
\]
Solving the marginal conditions, we obtain $\mu_1 = 1$, $\mu_2 = \delta(1 - \lambda)$, $\Delta q^* = \alpha(1 - \rho)(1 - \lambda) - \beta q_1$, $p_1 = \theta q_1$, and $p_2 = \alpha(1 - \lambda)(1 - \rho) - \beta q_1$. The resulting firm profits are:

$$n_{\text{firm}}^* = \theta q_1 + \frac{\alpha(1 - \lambda)(1 - \rho) - \beta q_1}{2\alpha}$$

**(High-end Focus and then Mass Market)**

Substituting $\{l_0 = 1, l_1 = 1\} \rightarrow D_1 = 1 - \lambda$, $\{l_1 = 1, l_0 = 1\} \rightarrow D_2 = 1$ in Eqs. (6) and (11), we obtain the Lagrangian

$$L = p_1(1 - \lambda) + \delta(p_2 - \rho \Delta q) - \frac{\alpha(1 - \rho) - \beta q_1}{2\alpha} - \mu_1(p_1 - q_1) - \mu_2(p_2 - \Delta q)$$

where $\mu_1 \geq 0$ and $\mu_2 \geq 0$.

**Second Period Price**

<table>
<thead>
<tr>
<th>$p_2$</th>
<th>$q_1$ + $\Delta q$</th>
<th>No sales to low-type consumers in $t-2$</th>
<th>Market Inversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta(q_1 + \Delta q)$</td>
<td>$\Delta q$</td>
<td>Market Penetration</td>
<td>High-end Focus</td>
</tr>
<tr>
<td>$\theta \Delta q$</td>
<td>$\theta q_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_1$</td>
<td>$q_1$</td>
<td>No sales in $t-2$</td>
<td>No market</td>
</tr>
</tbody>
</table>

**Fig. 3.** The characterization of market outcomes as a function of prices when consumers are myopic.

The Kuhn–Tucker conditions are

<table>
<thead>
<tr>
<th>Marginal condition</th>
<th>Complementary slackness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\delta q}{\delta q} - 1 - \lambda - \mu_1 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} p_1 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \delta - \mu_2 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} p_2 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \theta(1 - \lambda) + \theta q_1 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} \Delta q - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - q_1 - p_1 \geq 0$</td>
<td>$\frac{\delta q}{\delta q} \mu_1 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \Delta q - p_2 \geq 0$</td>
<td>$\frac{\delta q}{\delta q} \mu_2 - 0$</td>
</tr>
</tbody>
</table>

Solving the marginal conditions, we obtain $\mu_1 = 1 - \lambda$, $\mu_2 = \delta$, $\Delta q^* = \alpha(1 - \rho) - \beta q_1$, $p_1 = q_1$, and $p_2 = \alpha(1 - \lambda)(1 - \rho) - \beta q_1$. The resulting profits are:

$$n_{\text{firm}}^* = (1 - \lambda)(1 - \rho)q_1 + \frac{\alpha(1 - \rho) - \beta q_1}{2\alpha}$$

**(Market Inversion)**

Substituting $\{l_0 = 0, l_1 = 1\} \rightarrow D_1 = 1 - \lambda$, $\{l_1 = 0, l_0 = 1\} \rightarrow D_2 = \lambda$ in Eqs. (6) and (11), we obtain the Lagrangian

$$L = p_1(1 - \lambda) + \delta(p_2 - \rho \Delta q) - \frac{\alpha(\theta - \rho) - \beta q_1}{2\alpha} - \mu_1(p_1 - q_1) - \mu_2[p_2 - \theta(q_1 + \Delta q)]$$

where $\mu_1 \geq 0$ and $\mu_2 \geq 0$.

The Kuhn–Tucker conditions are

<table>
<thead>
<tr>
<th>Marginal condition</th>
<th>Complementary slackness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\delta q}{\delta q} - 1 - \lambda \theta q_1 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} p_1 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \delta - \mu_2 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} p_2 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \theta \Delta q - \theta q_1 \leq 0$</td>
<td>$\frac{\delta q}{\delta q} \Delta q - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - q_1 - p_1 \geq 0$</td>
<td>$\frac{\delta q}{\delta q} \mu_1 - 0$</td>
</tr>
<tr>
<td>$\frac{\delta q}{\delta q} - \Delta q - p_2 \geq 0$</td>
<td>$\frac{\delta q}{\delta q} \mu_2 - 0$</td>
</tr>
</tbody>
</table>

Solving the marginal conditions, we obtain $\mu_1 = 1 / \theta$, $\mu_2 = \delta / \theta$, $\Delta q^* = \alpha(\theta - \rho) - \beta q_1$, $p_1 = \theta q_1$, and $p_2 = \theta(\alpha(\theta - \rho) - \beta q_1)$. The resulting profits are:

$$n_{\text{firm}}^* = \theta q_1 + \frac{\alpha(\theta - \rho) - \beta q_1}{2\alpha}$$
We now identify the firm strategy that results in the market outcome yielding maximum profits. As before, comparing profits above we can see that \( \pi_{\text{HP}} > \pi_{\text{LP}} > \pi_{\text{AM}} \) due to Eq. (1). Comparing \( \pi_{\text{HP}} \) and \( \pi_{\text{AM}} \) we can see that \( \pi_{\text{HP}} \geq \pi_{\text{AM}} \) only if \( \alpha \geq \alpha_{\text{AM}} \) where

\[
\alpha_{\text{AM}} = \frac{2q_{1}}{1 - \rho + \lambda(\theta - \rho)} \left( \frac{\lambda \theta}{1 - \rho - \lambda(\theta - \rho)} + \beta \right). \tag{A.5}
\]

Therefore, High-end focus and then mass market is the optimal strategy of the firm if \( \alpha \geq \alpha_{\text{AM}} \). However, if \( \alpha < \alpha_{\text{AM}} \), Market inversion is the optimal strategy, Q.E.D.

**Proof of Proposition 2.** The equilibrium concept: To solve the game, we use Perfect Bayesian Equilibrium (PBE) concept. The PBE leads to a unique outcome when:

(A) The strategies of the informed players are optimal given the beliefs of the uninformed players.

(B) The beliefs of uninformed players are based on strategies that are consistent with Bayes' Rule.

The PBE imposes a rule of "logical consistency" on the beliefs of uninformed players ([Fudenberg & Tirole, 1991]); that is, the beliefs of uninformed players (i.e., consumers) are derived using Bayes' Rule from the actions of the informed player (the firm) before the uninformed player makes a decision. We assume that \( \mu \in [0,1] \) is the consumers' prior belief that the firm has a high PDC, having observed the firm action or signal \( p_{\text{L}} \) and its posterior belief. The triplet \( \{p^{H}_{\text{L}}, p^{H}_{\text{L}}, \mu\} \) constitutes a Perfect Bayesian Equilibrium (PBE) if and only if it satisfies the following conditions that are related to sequential rationality (P) and Bayesian consistency in beliefs (B).

(P) \( P^{H}_{\text{L}} = \arg \max p_{1}(\mu(p_{\text{L}})) \)

(B) If \( p^{H}_{\text{L}} = p^{H}_{\text{L}} = p_{\text{L}} \) then \( \mu(p^{H}_{\text{L}}) = \mu \). (Pooling Equilibrium)

If \( p^{H}_{\text{L}} \neq p^{H}_{\text{L}} \) then \( \mu(p^{H}_{\text{L}}) = 1 \) and \( \mu(p^{H}_{\text{L}}) = 0 \). (Separating Equilibrium)

In this game there are two possible types of equilibria. In a pooling equilibrium, consumers cannot update their prior belief by observing only the price because both high and low type firms charge the same price. Conversely, in a separating equilibrium, consumers can identify the firm type because the two types of firms charge different prices. PBE only imposes logical consistency on the beliefs of the players over actions on the equilibrium path; there are no restrictions on the beliefs of the players over actions off the equilibrium path. In signaling games, freedom in specifying off-equilibrium beliefs can lead to multiple equilibria when the off-equilibrium beliefs of uninformed players attribute positive probability to the informed player (the firm) choosing an equilibrium-dominated strategy. The Intuitive Criterion (IC) of Cho and Kreps (1987) eliminates these equilibria by imposing a restriction on the players' beliefs over actions off the equilibrium path.

**Intuitive Perfect Bayesian Equilibrium.**

A PBE violates the intuitive criterion if there exists an action that yields strictly greater payoffs for a player given that the uninformed players ascribe zero probability to a player's action that is “equilibrium-dominated.” An action is “equilibrium-dominated” for a player if that action leads to lower profits than another putative equilibrium. In other words, a firm type choosing an “equilibrium-dominated” action cannot increase its profit over what it earns under equilibrium.

In the context of this model, the beliefs of consumers subject to the IC restrict the high PDC firm to a set of strategies \( p^{H}_{\text{L}} \) which is equilibrium dominated for a firm with a low PDC, were it to implement a strategy from this set, the low PDC firm would earn less than its “guaranteed” level of profit. The only equilibrium that survives the intuitive criterion is a separating equilibrium with minimal inefficient signaling. In addition, a high PDC firm has a profit-increasing deviation from all possible pooling equilibria when signaling is possible.15 When a signal is either costless or inexpensive, signaling may be impossible.

The guaranteed profit for the low PDC firm, \( \pi_{L} \) is the profit it earns when consumers can observe quality. The low PDC firm has an incentive to mimic a high PDC firm if it increases profit by offering \( p^{H}_{\text{L}} \); the offer that would be made by a firm with a high PDC, i.e., \( \pi_{L}(p^{H}_{\text{L}}, \mu = 1) > \pi_{L} \). In equilibrium, the intuitive criterion rules out these strategies for the high PDC firm. As noted earlier, the intuitive criterion is the basis for the following constraint in the high PDC firm’s optimization:

\[
\pi_{L}(p^{H}_{\text{L}}, \mu = 1) > \pi_{L}. \tag{A.6}
\]

Note that a high PDC firm should not be able increase profits by pretending to be a firm with a low PDC, i.e.,

\[
\pi_{L}(p^{H}_{\text{L}}, \mu = 0) < \pi_{L}. \tag{A.7}
\]

Because \( q_{L} < q_{H} \), this restriction is satisfied.

The “guaranteed” profit of the low PDC firm is the profit it earns when it offers the price \( p_{1} \) which is the optimal price of the low PDC when quality is observable. The guaranteed profit, \( \pi_{L} \), is obtained by substituting \( \alpha = \alpha_{L} \) and \( q_{1} = q_{L} \) into the firm’s profit function and optimizing.

All putative pooling equilibria are unstable based on the intuitive criterion. Recall from Proposition 1 that the optimal strategy for the high (low) PDC firm is High-end focus and then mass market (Market inversion). When quality is unobservable, if the high PDC firm offers the same price which is also the maximum price it can charge as under complete information \( p_{1}(\theta_{1} = 1, q_{1} = q_{H}) = q_{H} \), it would violate the “no-mimic condition” (Eq. (A.6)). If a high PDC firm violates the no-mimic condition, a firm with a low PDC will have an incentive to offer the same price as that of the high PDC firm, which would result in a pooling equilibrium. We show below, using the Intuitive Criterion, that a pooling equilibrium does not exist.

Suppose there is a putative pooling equilibrium \( (p_{1}) \) in which the consumers accept the first period price given the expected quality \( E(q_{1}) = \mu_{q}\mu_{1} + (1 - \mu_{q})q_{L} \) given \( p_{1} \leq E(q_{1}) \).

**Step 1** First, we can find a deviation combination \( (p^{D} = p_{1} + D) \) such that

\[
\pi^{D}_{L}(p^{D}_{1}, \mu = 1) = \pi_{L}(p^{D}_{1}, \mu)
\]

where \( \pi^{D}_{L}(p^{D}_{1}, \mu = 1) = \pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) = \pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) \) and

\[
\pi^{D}_{L}(p_{1}, \mu = 1) = \pi^{D}_{L}(p_{1}, q_{1} = q_{L}, \alpha = \alpha_{L}). \tag{A.8}
\]

From the optimization in the Proof of Proposition 1, we can see that

\[
\pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) = (1 - \lambda)(p^{H}_{1} + D) + \lambda \pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) = (1 - \lambda)(p^{H}_{1} + D) + \lambda \pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) = \pi^{D}_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}).
\]

where \( D \) solves \( \pi_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) = (1 - \lambda)(p^{H}_{1} + D) + \lambda \pi_{L}(p^{H}_{1} = p_{L}, q_{1} = q_{L}, \alpha = \alpha_{L}) \). Now, considering a deviation price \( p^{D}_{1} \) which is infinitesimally less profitable than \( p^{H}_{1} \), i.e., \( p^{D}_{1} = p^{H}_{1} + D - \epsilon \), we have the low PDC firm's profit \( \pi^{D}_{L} < \pi^{H}_{L} \). Thus at the price \( D^{-} \), the low PDC firm earns strictly less profit than the equilibrium profit \( \pi^{H}_{L}(p^{H}_{1}, \mu) \), implying the price \( D^{-} \) is equilibrium dominated for the low PDC firm. According to the Intuitive Criterion, consumers cannot ascribe positive probability to a firm type choosing a strategy that is equilibrium dominated. Therefore, the posterior probability of the consumers \( \mu(D^{-}) = 1 \).

---

15 When a signal is either costless or inexpensive, signaling may be impossible.
Step 2 Using this price \( p_i^H \), the profit of the high PDC firm is \( \pi_H^0 = \pi_H[p_i^H; q_i^H; q_H; \alpha] = \alpha q_H \), \( \frac{\partial}{\partial q_H} [\pi_H[p_i^H; q_H]] \) (because \( \alpha_H > \alpha_M \) \( q_H > q_M \)). Therefore if the high PDC firm offers a price infinitesimally lower, i.e., \( p_i^H - \delta \), this price will still yield a positive profit \( \pi_H^0 > \pi_H[p_i^H; q_H] \). Thus the price \( p_i^H \) is not equilibrium dominated for the high PDC firm. The high PDC firm can increase its profits by offering a deviation price \( p_i^H \), and convince consumers that it has a high PDC and also earn a higher profit. Thus, there can be no intuitive pooling equilibrium.

Separating equilibrium for myopic consumers

We consider a separating equilibrium when quality is unobservable. We first obtain the no-mimic condition from the optimization problem of the low PDC firm when \( \alpha_L < \alpha_M \), using the separating price \( p_i^H \) as given below.\(^{20}\) Solving the game backwards, its second period pricing decision is

\[
\pi_L \left( p_i^H, \mu = 1 \right) = \max_{p_i^H} \pi_L \left( p_i^H \right) \text{ s.t. } p_i^L \leq \theta(q_L + \Delta q).
\]

The Lagrangian is given by

\[ L = (1-\lambda)p_i^H + \partial(q_L - \rho \Delta q) - \frac{\Delta a^2 + 2 \beta q_i^H \Delta q}{2 \alpha \theta} - \mu_i(p_i^H - \theta(q_L + \Delta q)) \]

where \( \mu_i \geq 0 \).

Kuhn-Tucker conditions yield \( \mu_i = \alpha_L \lambda > 0 \), \( \Delta q = \alpha_L \lambda (\theta - \rho) - \beta q_i^L \), and \( p_i^L = \theta(q_L + \Delta q) \) with resulting profits:

\[
\pi_L \left( p_i^H; \mu = 1 \right) = (1-\lambda)p_i^H + \alpha_L \theta q_i^L + \beta \frac{(\alpha_L \lambda (\theta - \rho) - \beta q_i^L)^2}{2 \alpha_L}.
\] (A.9)

Note that the low PDC firm mimics only if \( \pi_L \left( p_i^H; \mu = 1 \right) > \pi_L^0 \), which is possible only if \( p_i^H > q_i \), where \( p_i^H \in [q_i, \mu_i] \). The no-mimic constraint is obtained by simplifying the condition \( \pi_L \left( p_i^H; \mu = 1 \right) \leq \pi_L^0 \) where \( \pi_L^0 \) is obtained from the expression (A.3) as given by

\[
\pi_L^0 = |(1-\lambda)(1-\theta)| q_i + \delta \frac{(\alpha_L \lambda (\theta - \rho) - \beta q_i^L)^2}{2 \alpha_L}.
\] (A.10)

The no-mimic constraint therefore is given by

\[
\partial p_i^H \leq q_i.
\] (A.11)

We solve the separating profit maximization problem of the high PDC \( (\alpha_H > \alpha_M(q_L = q_M)) \) firm using the "no-mimic condition" (Eq. (A.11)) as a constraint: \( p_i^H \leq q_i \). This is given by the optimization problem

\[
\pi_H^0 \left( p_i^H; \mu = 1 \right) = \max_{p_i^H} \pi_H^0(\alpha = \alpha_H, q_i = q_M) \text{ s.t. } p_i^H \leq q_i.
\]

Simplifying the Lagrangian only for the decisions \( p_i^H \) (note that the decisions \( p_i^L \) and \( \Delta q \) remain unchanged from the complete information case), we have

\[ L = (1-\lambda)p_i^H + \frac{\alpha_H (1-\rho) - \beta q_i^H^2}{2 \alpha_H} - \mu_i(p_i^H - q_i) \]

where \( \mu_i \geq 0 \).

\(20\) Note that because \( \alpha_L < \alpha_M(q_L = q_M) \), Market Inversion is the optimal strategy for the firm.

The Kuhn-Tucker conditions yield: \( \mu_1 = 1 - \lambda > 0 \) and \( p_i^H = q_i \). Note that this may lead to a tie, which is broken by the high PDC firm by charging \( p_i^H = q_i - \varepsilon \) where \( \varepsilon > 0 \) is infinitesimally small. The separating profit for the high PDC firm therefore is

\[
\pi_H^0 \left( p_i^H, \mu = 1 \right) = (1-\lambda)q_i + \delta \frac{(\alpha_H (1-\rho) - \beta q_i^H)^2}{2 \alpha_H}.
\]

Q.E.D.

Proof of Proposition 3. If the quality of the high PDC firm is above a threshold given by the following condition, the Lows buy in period 1

\[ p_i^H \leq \theta q_i \Leftrightarrow q_i \geq q_i/0. \] (A.12)

They do not, however, trade up in period 2, as the market penetration strategy is preferable to the mass market strategy, i.e., \( \pi_{HFP} > \pi_{MM} \) from Eqs. (A.13) and (A.16) under Proof of Proposition 1. Q.E.D.

Proof of Proposition 4. The solution approach is exactly the same as that in Proof of Proposition 1. Fig. 4 illustrates a characterization of the market outcomes as a function of the prices of the two periods. We apply the Kuhn-Tucker necessary and sufficient conditions for global maximum for each of the four strategies below as the objective function is concave and the constraints are linear.

Market penetration. Substituting \( I_0 = I_1 = 1 \Rightarrow D_1 = 1 \), and \( I_1 = 1, I_0 = 0 \Rightarrow D_2 = 1, \lambda \), in Eqs. (6) and (15), we obtain the Lagrangian

\[ L = p_i + \delta(p_i - \rho \Delta q)/[1-\lambda] - \delta \frac{\Delta a^2 + 2 \beta q_i^H \Delta q}{2 \alpha} - \mu_i(p_i - \theta q_i + \theta (\Delta q - p_i)) - \mu_2(p_2 - \Delta q) \]

where \( \mu_1 \geq 0 \) and \( \mu_2 \geq 0 \).

The Kuhn-Tucker conditions are

\[
\begin{align*}
\frac{\partial}{\partial p_i} & = 1 - \lambda \leq 0 \\
\frac{\partial}{\partial \Delta q} & = 0 \\
\frac{\partial}{\partial q_i} & = \alpha(q_i - q_M) \leq 0 \quad q_i \geq 0.
\end{align*}
\]

Solving the marginal conditions, we obtain \( \mu_1 = 1, \mu_2 = \delta(2-\lambda), \Delta q^* = \alpha q_i - \theta (1-\lambda)q_i - \theta q_i, \quad p_i = \theta q_i + \delta (1-\lambda) (1-\lambda)q_i - \theta q_i, \) and \( p_i = \alpha q_i - \theta (1-\lambda)q_i - \theta q_i, \) the resulting firm profits are:

\[
\pi_{HFP}^0 = \theta q_i + \delta \frac{\alpha(1-\theta)(1-\lambda) - \beta q_i^H}{2 \alpha}.
\] (A.13)

High-end focus and then mass market. Substituting \( I_0 = 0, I_1 = 1 \Rightarrow D_1 = 1, \lambda \), \( I_1 = 1, I_0 = 0 \Rightarrow D_2 = 1, \) in Eqs. (6) and (16), we obtain the Lagrangian

\[
\begin{align*}
L & = p_i (1 - \lambda) + \delta(p_i - \rho \Delta q) - \delta \frac{\Delta a^2 + 2 \beta q_i^H \Delta q}{2 \alpha} - \mu_i(p_i - q_i + \theta (\Delta q - p_i)) - \mu_2(p_2 - \Delta q)
\end{align*}
\]

where \( \mu_1 \geq 0 \) and \( \mu_2 \geq 0 \).
Second Period Price

\( p_2 \)

<table>
<thead>
<tr>
<th>( q_1 + \Delta q )</th>
<th>( \theta(q_1 + \Delta q) )</th>
<th>( \Delta q )</th>
<th>( \theta \Delta q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No sales in ( t=1 ) &amp; Market Inversion</td>
<td>No sales in ( t=2 ) &amp; No market</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. The characterization of market outcomes as a function of prices when consumers are forward-looking.

The Kuhn–Tucker conditions are

<table>
<thead>
<tr>
<th>Marginal condition</th>
<th>Complementary slackness</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\partial}{\partial \mu} - 1 - \lambda - \mu_1 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu_1} p_1 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu_2} - \delta + \delta(1 - \lambda) - \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu_2} p_2 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{p}_1 - \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \Delta q = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{p}_1 - \delta \hat{\Delta q} - \delta \hat{\Delta q} - \delta \hat{q}_1 + \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \mu_1 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{q}_1 - \delta \hat{\Delta q} - \delta \hat{p}_2 - \mu_2 \geq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \mu_2 = 0 )</td>
</tr>
</tbody>
</table>

Solving the marginal conditions, we obtain \( \mu_1 = 1 - \lambda, \mu_2 = \delta \). Note, here, that from the marginal condition \( \Delta \leq 0 \). \( \delta \alpha(\theta - \rho - (1 - \lambda)) - \beta q_1 - \gamma \leq 0 \) under condition 1. Therefore, from the complementary slackness, we obtain \( \Delta q^* = 0 \). From the rest of the marginal conditions, we obtain \( p_1 = (1 + \delta)q_1 \), and \( p_2 = \theta q_1 \). Note, however, that because \( \Delta q^* = 0 \), the forward-looking firms will anticipate the subsequent drop in price to \( \theta q_1 \) in period 2, and therefore, defer their purchase (Coase, 1972), forcing the firm to charge \( p_1 = \theta q_1 \) and leading to profits:

\[
\pi_{\text{firm}} = \theta q_1. \tag{A.15}
\]

Mass market. Substituting \( \{I_0 = I_1 = 1 \} \rightarrow D_1 = 1, \{J_1 = 1 \} \rightarrow D_2 = 1 \) in Eqs. (6) and (17), we obtain the Lagrangian

\[
L = p_1(1 - \lambda) + \delta(p_2 - \rho \Delta q) - \delta \Delta q^2 + 2[\delta q_1 \Delta q - \mu_1(p_1 - \hat{\Delta q} - \hat{p}_2) + \mu_2(p_2 - \theta q_1 + \Delta q)]
\]

where \( \mu_1 \geq 0 \) and \( \mu_2 \geq 0 \).

The Kuhn–Tucker conditions are

<table>
<thead>
<tr>
<th>Marginal condition</th>
<th>Complementary slackness</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\partial}{\partial \mu} - 1 - \lambda - \mu_1 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu_1} p_1 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu_2} - \delta + \delta(1 - \lambda) - \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu_2} p_2 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{p}_1 - \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \Delta q = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{p}_1 - \delta \hat{\Delta q} - \delta \hat{q}_1 + \mu_2 \leq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \mu_1 = 0 )</td>
</tr>
<tr>
<td>( \frac{\partial}{\partial \mu} - \delta \hat{q}_1 - \delta \hat{\Delta q} - \delta \hat{p}_2 - \mu_2 \geq 0 )</td>
<td>( \frac{\partial}{\partial \mu} \mu_2 = 0 )</td>
</tr>
</tbody>
</table>

Solving the marginal conditions, we obtain \( \mu_1 = 1 - \lambda, \mu_2 = 2 \delta, \Delta q^* = \delta \alpha(\theta - \rho - (1 - \lambda)) - \beta q_1 - \gamma \leq 0 \) under condition 1. Therefore, from the complementary slackness, we obtain \( \Delta \leq 0 \). \( p_1 = (1 + \theta)q_1, \) and \( p_2 = \theta q_1 \). The resulting profits are:

\[
\pi_{\text{firm}} = \theta q_1 + \frac{\delta(\alpha(\theta - \rho) - \beta q_1)^2}{2\alpha}. \tag{A.16}
\]

We now identify the firm strategy that results in the market outcome yielding maximum profits. Comparing profits above, we
can see that $\pi_{LP} > \pi_{LM} > \pi_{LM}$ and $\pi_{LP} > \pi_{LM}$ due to Eq. (1). Comparing $\pi_{LP}$ and $\pi_{LM}$, we can see that $\pi_{LP} \geq \pi_{LM}$ Next comparing $\pi_{LP}$ with $\pi_{LM}$, we can see that $\pi_{LP} \geq \pi_{LM}$ only if $\alpha \leq \alpha_1$, where

$$\alpha_1 = 2q_1 \left( \frac{1 - \gamma}{\bar{\rho}} \right) + \beta \left( \gamma - (1 - \gamma)(1 - \rho) \right).$$

(A.17)

Therefore, Market penetration is the optimal strategy of the firm if $\alpha \leq \alpha_1$. However, $\pi_{LP} > \pi_{LM}$ if $\alpha > \alpha_1$, i.e., High-end focus and then mass market is the optimal strategy. Q.E.D.

Proof of Proposition 5. The equilibrium concept: Is the same as that which was stated under Proof of Proposition 2.

All putative pooling equilibria are unstable based on the intuitive criterion

Analogous to the case of myopic consumers (see Proof of Proposition 2), the proof that all putative pooling equilibria are unstable follows the same logic. Recall from Proposition 4 that the optimal strategy for the low (L) PDC firm is Market penetration (High-end focus and then mass market). If the high PDC firm offers the same price as under complete information, i.e., $p_j = (1 - \lambda)(\lambda - \Delta)$, it would violate the “no-mimic condition” (Eq. (A.6)). Therefore if the high PDC firm will have an incentive to offer the same price, which would result in a pooling equilibrium. We show below, using the Intuitive Criterion, that a pooling equilibrium does not exist.

Suppose there is a putative pooling equilibrium $(p_j^*)$ in which the consumers accept the first period price given the expected quality $E(q_1) = \mu q_1 + (1 - \mu) q_2$ given $p_j^* \leq E(q_1)$. Step 1 First, we can find a deviation combination $(p_j^1 = p_j^* + D)$ such that

$$\Delta \pi_j^1(p_j^1, \mu^* = 1) = \Delta \pi_j^1(p_j^*, \mu)$$

where $\Delta \pi_j^1(p_j^1, \mu^* = 1) = \pi_{LP}^1(p_j^1, \mu^* = 1)$ and $\Delta \pi_j^1(p_j^1, \mu) = \pi_{LP}^1(p_j^1, \mu)$.

(A.18)

From the optimization in the Proof of Proposition 4, we can see that $\pi_{LP}^1(p_j^1, \mu^* = 1) = \pi_{LP}^1(p_j^1, \mu) = \pi_{LP}^1 + D$ + $\frac{\delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \mu}{\alpha_j^2}$ and $\pi_{LP}^1(p_j^1, \mu^* = 1)$ solves $\pi_{LP}^1(p_j^1, \mu^* = 1) = \pi_{LP}^1 + D - \epsilon$, we have the low PDC firm's profit $n_{LP}^1 - n_{LP}^1$ as $\epsilon \rightarrow 0$. Thus at the price $D$, the low PDC firm earns strictly less profit than the equilibrium profit $\pi_{LP}^1(p_j^1, \mu)$, implying the price $D$ is equilibrium dominated for the low PDC firm. According to the Intuitive Criterion, consumers cannot ascribe positive probability to a firm type choosing a strategy that is equilibrium dominated. Therefore, the posterior probability of the consumers $\mu(D) = 0$.

Step 2 Using this price $(p_j^1)$, the profit of the high PDC firm is $\pi_{LP}^1 = \pi_{LP}^1(p_j^1, \mu) = \pi_{LP}^1 + D$ + $\frac{\delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \mu}{\alpha_j^2}$ (because $\alpha_j > \alpha_j$ (q_1 = q_2) > $\alpha_j$ q_1 = q_2) > $\alpha_j$ q_1 = q_2). Therefore if the high PDC firm offers a price infinitesimally lower, i.e., $p_j^1$, this price will still yield a positive profit $\pi_{LP}^1 > \pi_{LP}^1(p_j^1, \mu)$. Thus the price $p_j^1$ is not equilibrium dominated for the high PDC firm. The high PDC firm can increase its profits by offering a deviation price $p_j^1$ and convince the consumers that it has a high PDC and also earn a higher profit. Thus, there can be no intuitive pooling equilibrium. Q.E.D.

Separating equilibrium for forward-looking consumers

To identify the separating equilibrium, we first obtain the no-mimic condition from the optimization problem of the low PDC firm using the separating price $p_j^1$. Solving the game backwards, period 2 pricing decision is

$$\pi_j^1(p_j^1, \mu^* = 1) = \pi_j^1(p_j^1, \mu) + 2 \pi_j^1(p_j^1, \mu^* = 1) \Delta q_2^1$$

The Lagrangian is given by

$$L = p_j^1 + \delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \Delta q_2^1$$

$$= (1 - \lambda) - \beta q_j^1 \Delta q_2^1$$

and $p_j^1 = \Delta q_2^1 \mu_2 \Delta q$ with resulting profits:

$$\pi_j^1(p_j^1, \mu^* = 1) = p_j^1 + \delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \Delta q_2^1 \mu_2 \Delta q$$

Note that the low PDC firm mimics only if $\pi_j(p_j^1, \mu^* = 1) > \pi_j^1$ where

$$\pi_j(p_j^1, \mu^* = 1) = \pi_j(p_j^1, \mu^* = 1) + \delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \Delta q_2^1 \mu_2 \Delta q$$

This condition simplifies to $p_j^1 > p_j^1$ where

$$\pi_j^1(p_j^1, \mu^* = 1) = \pi_j(p_j^1, \mu^* = 1) + \delta \alpha_j(1 - \lambda)(1 - \rho) - \beta q_j^1 \Delta q_2^1 \mu_2 \Delta q$$

Note from the optimization problem for Market penetration (Eq. (A.13)) that $p_j^1 < p_j^1$ where $p_j^1$ is the complete information price of the high PDC firm) is a necessary condition for mimicking. Otherwise, the high PDC firm can separate using its complete information price $p_j^1$. The no-mimic constraint therefore is given by $p_j^1 < p_j^1$. Solving the separating profit maximization problem of the high PDC firm involves the “no-mimic condition” (Eq. (A.21)) as a constraint. This is given by the optimization problem

$$\pi_j^1(p_j^1, \mu^* = 1) = \max \pi_j(p_j^1, \mu = \alpha_j) = \pi_j(p_j^1, \mu = \alpha_j)$$

Simplifying the Lagrangian only for the decisions $p_j^1$ (note that the decisions $p_j^1$ and $\Delta q_2^1$ remain unchanged from the complete information case) and solving the marginal conditions, we obtain: $\mu_1 = 1 - \lambda \geq 0$ and $p_j^1 = p_j^1$. The separating profit for the high PDC firm therefore is

$$\pi_j^1(p_j^1, \mu^* = 1) = p_j^1 + \delta \alpha_j(1 - \lambda(1 - \rho) - \beta q_j^1 \Delta q_2^1 \mu_2 \Delta q$$

Note that both high and the lows buy in period 1 and hence Market penetration is the optimal strategy also under signaling. Q.E.D.

References


An introduction to the application of (case 1) best–worst scaling in marketing research

Jordan Louviere a,⁎, Ian Lings b, Towhidul Islam c, Siegfried Gudergan d, Terry Flynn a

a Centre for the Study of Choice (CenSoC), University of Technology, Sydney, PO Box 123 Broadway, NSW 2007, Australia
b School of Advertising, Marketing and Public Relations, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia
c Department of Marketing and Consumer Studies, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada
d The University of Newcastle, Social Sciences Building, University Drive, Callaghan, NSW 2308, Australia

Abstract

We review and discuss recent developments in best–worst scaling (BWS) that allow researchers to measure items or objects on measurement scales with known properties. We note that BWS has some distinct advantages compared with other measurement approaches, such as category rating scales or paired comparisons. We demonstrate how to use BWS to measure subjective quantities in two different empirical examples. One of these measures preferences for weekend getaways and requires comparing relatively few objects; a second measures academics’ perceptions of the quality of academic marketing journals and requires comparing a significantly large set of objects. We conclude by discussing some limitations and future research opportunities related to BWS.

1. Introduction

Academics and practitioners in various disciplines often wish to measure an individual’s strength of preference for (or level of agreement with) a number of objects (which can be statements or some other item of interest). A typical objective is to locate all the objects on a measurement scale with known mathematical properties to allow robust statistical comparisons of changes over time and/or differences among respondents. In practice, this can be challenging. For example, rating scales attempt to ensure that all individuals use the same numerical scale, but in practice, various idiosyncrasies in response styles have been found (Auger, Devinney, & Louviere, 2007). Such idiosyncrasies can arise from individuals using rating scales in different ways, from cultural differences and/or from verbal ambiguities with labels (Lee, Soutar, & Louviere, 2008). Furthermore, it has been observed that individuals tend not to discriminate between response categories when they are not asked to respond in ways that elicit tradeoffs or relative preferences for the objects being valued, such as asking people to rate the “importance” of several factors on a rating scale. That is, respondents do not have to trade off one factor against another, as evidence indicates that this often leads to minimal differences in mean ratings (e.g., Cohen & Neira 2003; Lee, Soutar, & Louviere, 2007).

An approach to dealing with such issues that has been growing in popularity in many fields is to avoid tasks that ask individuals to use numbers in favor of tasks that infer strength of preference (or other subjective, latent dimensions) from how often they choose one object over another, known objects. Such observed choice frequencies ensure that the derived numbers are on a known (choice frequency or probability) scale. However, some choice-based approaches, such as the method of paired comparisons, require large numbers of choice questions to estimate preferences for objects. Indeed, asking individuals to choose from all possible pairs of objects is not feasible in survey settings as the number of objects grows, a clear weakness of the method of paired comparisons.

The purpose of this paper is to introduce, discuss and illustrate a choice-based measurement approach that reconciles the need for question parsimony with the advantage of choice tasks that force individuals to make choices (as in real life). Prior work recognizes three choice-based measurement cases. In case 1 (the object case), individuals are asked to choose the best and worst (on some subjective scale) from a set of objects (e.g., Finn & Louviere, 1992). In case 2 (the profile case), individuals evaluate several profiles of objects described by combinations of attributes/features dictated by an underlying design; they “see” the profiles one at a time and choose the best and worst feature/attribute levels within each presented profile (e.g., Louviere, 1994). In case 3, individuals choose the best and the worst designed profiles (choice alternatives) from various choice sets dictated by an underlying design (e.g., Marley & Pihlens, 2012).

The purpose of this paper is to introduce, discuss and illustrate case 1. We focus on case 1 because it illustrates the fundamentals of choice-based measurement in general and what is known as “best–worst scaling” (BWS) in particular. BWS was introduced by Finn and Louviere (1992), and recent advances suggest that academics and practitioners would benefit from an updated discussion of its concepts and methods.
BWS is one way to avoid and overcome some of the limitations of rating-based and similar measurement methods used in marketing and in other fields. BWS case 1 typically allows one to obtain measures for each person (respondent) on a difference scale with known properties (Marley & Louviere, 2005). Cases 2 and 3 can be viewed as extensions of case 1 in which objects or items are represented as multi-dimensional choice objects (options). However, the fundamental ideas and principles from case 1 also apply to cases 2 and 3; thus, we focus on explaining case 1 in detail because this provides a foundation for understanding cases 2 and 3.

Accordingly, the objective of this paper is to provide an introduction for academics and practitioners on how to design, implement and analyze case 1 BWS studies. The case for such a paper is threefold. 1) As papers detailing the mathematical proofs of the main estimators used to implement such studies are highly technical and not easily understood by novices (Marley & Louviere, 2005; Marley, Flynn, & Louviere, 2008), there is a need for a more straightforward explanation to encourage applications. 2) Disciplines in which comprehensive ‘how to’ BWS discussions have been published have seen a proliferation of empirical studies (e.g., Flynn, 2010), suggesting that a tutorial paper should benefit marketing academics and practitioners. 3) Several methods for estimating the values of objects using underlying subjective scales have been proposed, but many of these, although easy to implement in a spreadsheet or generic statistical package, are not part of the typical ‘toolbox’ of methods used by academics and practitioners. Indeed, a ‘user guide’ paper detailing the BWS profile case (case 2) for health economists arose from requests to conferences (among other things) ‘see’ what the data and regression models ‘look’ like (Flynn, Louviere, Peters, & Coast, 2007).

Accordingly, to provide a ‘how to’ BWS tutorial, this paper is organized as follows. First, we offer a conceptual framework and empirical justification for BWS. We then present two empirical studies. We emphasize how to set up, design and implement a BWS case 1 survey in practice, and how to analyze the associated results. Specifically, we present worked examples that illustrate how to use BWS for relatively small (six objects) as well as very large (72 objects) comparison sets. The paper ends with a discussion and conclusion section that recaps the major points of the paper, identifies some limitations and issues and suggests some potential future research directions.

2. A conceptual framework for BWS

BWS is underpinned by random utility theory (RUT), which also underlies discrete choice experiments used in marketing research and economics (McFadden, 1974; Thurstone, 1927). RUT assumes that an individual’s relative preference for object A over object B is a function of the relative frequency with which A is chosen as better than, or preferred to, B. Thus, it requires individuals to make choices stochastically (with some error). Thurstone’s (1927) paper proposed RUT and used it to motivate and develop the method of paired comparisons, where individuals choose the ‘best’ object from sets of two objects. Thurstone recognized that the theory requires individuals to make errors in their choices, thus allowing the model parameter estimates that we term ‘scale values’ to be derived. Scale values are measures of the locations of each object on an underlying subjective scale of interest. McFadden (1974) generalized Thurstone’s RUT model to provide tractable, closed-form models that accommodate choices from sets of three or more objects. More formally, for the ‘best’ only case McFadden considered:

\[ S_A = V_A + \varepsilon_A \]
\[ S_B = V_B + \varepsilon_B \]
\[ S_C = V_C + \varepsilon_C \]
\[ S_D = V_D + \varepsilon_D. \]

In the above, the true subjective scale value \( S_k \) of the kth object consists of two components, the observed value \( V_k \), which is systematic (explainable), and the errors \( \varepsilon_k \), which are random (unexplainable). The random component implies that one cannot predict the exact choice that a person will make, but only the probability that a person will choose each object offered (McFadden, 1974). This choice probability can be expressed as:

\[ P(U = \text{best}(A, B, C, D)) = P[V_A + \varepsilon_A > V_k + \varepsilon_k]. \]

Considering that all other options are available to be chosen in the comparison set, McFadden (1974) derived what is known as the conditional logit model by assuming that the errors are distributed as independent and identically distributed Type 1 Extreme Value. The choice probabilities for this model have the following closed form expression:

\[ P(A = \text{best}(A, B, C, D)) = \exp(V_A)/[\exp(V_A) + \exp(V_B) + \exp(V_C) + \exp(V_D)]. \]

McFadden’s framework relates choices from sets of multiple objects to an underlying latent scale value associated with each object, but until recently, little work was available to help researchers identify and implement reasonably good ways of collecting choice data from individuals to implement these models. An obvious exception, of course, is the method of paired comparisons, which has been extensively studied (e.g., David, 1988). Unfortunately, the method of paired comparisons poses inherent limitations in survey applications because the number of comparisons needed increases geometrically with the number of objects to be measured. Thus, paired comparisons can be practical for measuring a few objects (e.g., six objects require 15 pairs), but typically are not practical for larger numbers of objects (e.g., we later study 72 objects, which would require 2556 pairs).

One way to address the size limitation of paired comparisons is the multiple choice approach introduced by Louviere and Woodworth (1983) that relies only on ‘best’ choices. Although their discrete choice experiment (“DCE”, also called “choice-based conjoint”) approach is widely used, few researchers seem to appreciate that collecting only “first (or best) choices” provides minimal information for statistical estimation purposes. Thus, an approach that provides more statistical information than merely the first or best choice could be useful in many research applications.

BWS capitalizes on the fact that collecting ‘worst’ information, in a similar way to ‘best’ information, provides much more information. That is, BWS capitalizes on the idea that when individuals evaluate a set of three or more objects or items on a subjective scale, their choices of the top and bottom objects/items should be (all else equal) more reliable than choices of middle objects/items. Thus, BWS assumes that individuals make reliable and valid choices of the two most extreme objects/items in a set, consistent with the adaptation level theory (Helson, 1964). A key advantage of BWS is that it provides information about both the top ranked and bottom ranked items in a set. Taken together, these two choices provide much more information about the ranking of the choice options in each set. Only order information matters in choices; hence, asking for both top and bottom ranked choices provides much more information about the overall ranking of the objects than just the top choice.

More generally, BWS implies use of multiple comparison sets, with each set having at least three objects/items. In this respect, a BWS “experiment” is just another type of DCE, similar to the DCEs proposed by Louviere and Woodworth (1983). To wit, they proposed constructing comparison (choice) sets from 2^J fractional factorial designs (J = the number of objects/items). However, most BWS applications design choice (comparison) sets with balanced incomplete block designs (BIBDs), such as Lee et al. (2008). A BIBD is a type of experimental design in which each choice option appears equally often, and co-appears equally often with each other choice option. Unlike 2^J designs, BIBDs ensure that choice set sizes are always equal.

A type of BIBD called a “Youden” design (e.g., Raghavarao, 1988) allows one to control for order by ensuring that each object appears in every order. In our experience, there is little difference in outcomes associated.
3. Implementing best–worst choice tasks: empirical example one

3.1. Empirical issue of interest

The first empirical example involves holiday destinations. The subjective dimension of interest is the likelihood of visiting a destination for a weekend getaway. The study population is residents of Sydney, Australia, who could choose among the following weekend getaways: 1) Central Coast (beach house), 2) Katoomba (up-market hotel), 3) Barrington Tops (isolated setting), 4) Bowral (Southern Highlands), 5) South Coast (heritage village), and 6) Sydney (up-market hotel). We recruited a sample of 420 respondents from the Pureprofile online panel in Australia who satisfied the criterion that they resided in the Sydney metropolitan areas (defined by postcodes) and had taken at least one weekend getaway holiday in the previous 12 months. The Pureprofile panel has over 600,000 households recruited and maintained that match, as closely as possible, the overall Australian population on key census demographics.

3.2. Implementing BWS tasks: design

The first stage in implementing a BWS survey is to choose a statistical design to construct the comparison sets. As noted earlier, researchers can choose from several statistical designs. We used a BIBD to design the comparison sets because it provides a constant comparison set sizes; b) increasing number of comparison sets approximately linearly in J (number of objects/items to be measured; here J = 6), such that one can often (but not always) find BIBDs for J objects/items in J or at most a few more than J sets; and c) BIBDs that can be found in many sources, such as Raghavarao (1988) and Street and Street (1987). BIBDs also ensure that each of the J objects/items occur the same number of times across all sets, and co-occur the same number of times with the other (J – 1) objects. These properties are important because unequal set sizes may unintentionally signal to individuals that a survey is about something unintended by the researcher and/or that they are “supposed to choose differently in sets of different sizes, etc.” (i.e., if set sizes differ across a survey, it may lead to “demand artifacts”). Additionally, if one object appears more often than other objects, it may signal that the survey is “really about” the one or more objects that appear more often.

To use a BIBD to implement a BWS survey, one numbers the objects/items of research interest from 1 to J and replaces the same (1 to J) numbers in a BIBD table with the corresponding names, symbols or descriptions of each object to be measured. We illustrate this below with a BIBD for six objects (coded 1, 2, …, 6) that creates 10 comparison sets (survey questions), as shown in the first 4 columns of Table 1.

Next, one uses a “find and replace” procedure to replace the object code numbers in columns 2 to 4 of Table 1 with the object names (here, holiday destinations, but more generally, they can be items, principles, brands, etc.) to create comparison sets, as shown in columns 5 to 7. The next step is to embed the comparison sets into a survey. One way to ask BWS questions is evidenced in the particular survey format, as displayed in Table 2.

<table>
<thead>
<tr>
<th>Set</th>
<th>Object codes</th>
<th>Object names</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 2 3 4 5 6</td>
<td>Central Coast beach house Katoomba upmarket hotel Barrington Tops, an isolated setting Bowral, Southern Highlands Sydney, upmarket hotel</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5 6</td>
<td>Central Coast beach house Katoomba upmarket hotel Barrington Tops, an isolated setting Bowral, Southern Highlands Sydney, upmarket hotel</td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 4 5 6</td>
<td>Central Coast beach house Katoomba upmarket hotel Barrington Tops, an isolated setting Bowral, Southern Highlands Sydney, upmarket hotel</td>
</tr>
<tr>
<td>4</td>
<td>1 2 3 4 5 6</td>
<td>Central Coast beach house Katoomba upmarket hotel Barrington Tops, an isolated setting Bowral, Southern Highlands Sydney, upmarket hotel</td>
</tr>
<tr>
<td>5</td>
<td>1 2 3 4 5 6</td>
<td>Central Coast beach house Katoomba upmarket hotel Barrington Tops, an isolated setting Bowral, Southern Highlands Sydney, upmarket hotel</td>
</tr>
</tbody>
</table>

Table 2 reveals that the BIBD used in the destination survey has the property that each destination occurs five times and co-occurs twice across the 10 sets.

3.3. BWS response data and simple analyses

A conventional BWS task requires respondents to choose the best and the worst objects in each comparison set. In this example, 420 respondents were asked to identify the destination s/he was most likely to visit (best) and which destination s/he was least likely to visit (worst) in each of the 10 comparison sets, each of which contained three destinations. The analysis is based on assigning the most likely destination a ‘+ 1’ and the least likely destination a ‘− 1’, and as each item appears five times, preferences are measured on a scale bounded by − 5 and + 5. A simple analysis involves summary statistics such as sums or means. Our survey also asked respondents to self-report the average number of trips they took to each destination during the previous 12 months (revealed preference), which allows us to compare BWS measures against these ‘revealed preference’ (RP) self-reports.

Columns 5 and 6 of Table 3 reflect one person’s best and worst choices. Summarizing the best and worst choice data merely involves counting the occurrences of best and worst choices for each choice object, as shown in Table 4. A simple scale and semi-order (ranking) is obtained by subtracting worst counts from best counts, as shown in the last column. The advantage of collecting data about the worst choice becomes clear when considering the lower ranked objects: the scale derived from best only choice data cannot distinguish objects 5 and 6, whereas best and worst choices, taken together, provide rank order information for these two objects. The reason for calculating counts is to obtain empirical estimates of the choice proportions. As noted by Louviere and Woodworth (1983), these counts contain all the statistical information in the data and can be used to estimate the parameters of a Luce (1959) or multinomial logit model (McFadden, 1974). That is, if there are J objects, one simply estimates the intercepts or “alternative-specific constants” in the logit model:
For aggregates of individuals

- One calculates the total number of times that each object of interest is chosen as best and worst across all comparison sets (choice sets) and individuals.
- If the data are disaggregated to represent each object in each comparison set for each person, the dependent variables resulting from this can be a) the difference in the best and worst counts or choice totals (i.e., best counts minus worst counts), b) a full or partial ranking obtained from the best and worst choices in each comparison set, or c) an expansion of the best and worst choices into implied choice sets, as discussed by Louviere et al. (2008).
- If the data are aggregated to represent each object across all comparison sets and persons, the dependent variables resulting from this can be a) the difference in the best and worst counts or choice totals (i.e., best counts minus worst counts) and/or b) the square root of the ratio of best counts divided by worst counts (this measure is proportional to the best counts under the assumption that worst counts = 1 / (best counts); the natural log of this quantity is the expected value of the object on the latent scale if the choice process is consistent with a conditional logit model).

The above dependent variables can be analyzed in relatively simple ways, such as calculating them directly from the data, or in sophisticated analytical ways, such as ordinal regression models and probabilistic discrete choice models. We show in Table 5 and Fig. 1 that ordinary least squares applied directly to the best minus worst differences will typically produce reliable and valid measures of the scale positions of the objects on the latent scale. We also note that the scale positions represent intercept terms in discrete choice models.

Researchers should be cautious when using BWS scores (best minus worst counts) to make inferences about individual respondents (Flynn, 2010). In the present example, objects 3 and 4 (in the middle of the scale) are not differentiated by BWS scores. Despite this caveat, experience suggests that one does not have to aggregate choices across many individuals for average scores (for individuals in question) to perform well and correlate highly with estimates from more sophisticated models.

3.4. BWS more sophisticated analyses requiring statistical software

We now consider two other ways to analyze the data, both of which utilize a [0,1] dependent variable (indicating whether a particular choice option was not chosen/chosen): 1) linear probability models (LPMs), which are ordinary least square (OLS) regression models fit to the choice data and 2) conditional logit models (McFadden, 1974). LPMs are justified by Heckman and Snyder (1997), who argue that if errors are not symmetric, LPMs are likely to better represent respondents’ choices than are other models (see also Alrich & Nelson, 1984).

To estimate LPMs and conditional logit models (CLMs), we use information from the best and worst choices in each comparison set to “expand the data”. By “expanding the data”, we mean using the rank order information to construct several sets of implied choices for various pairs of objects associated with each comparison set. For example, if one observes best and worst choices for two objects in a set of three objects, one has the full-rank ordering. One can use the observed best and worst differences and other differences to form several sets of implied choices.
worst choices to construct three pairs of implied choices (Horsky & Rao, 1984), from which one infers the implied choice that should be made for each of the three pairs of objects. Thus, for three option sets, one can “expand” best and worst choices into three implied choice pairs for each three option set (A vs B, B vs C and A vs C). For four object/option sets, best and worst choices give a semi-order; that is, one can infer the choices in most, but not all of the possible pairs. For other set sizes, the idea is the same, but observing only best and worst choices gives less information about a full ranking as the number of options per set increases.

More specifically, suppose one observes best and worst choices in a set of three options — say “A”, “B” and “C”. Suppose further that A is chosen best and C is chosen worst. This ranking implies that A should be chosen for the pair AB and the pair AC, and B should be chosen for the third pair BC. In the case of four options, say “A”, “B”, “C” and “D”, if a respondent chooses A as the best and D as the worst, this implies that A should be chosen in the pairs AB, AC and AD; B should be chosen in the pair BD; and C should be chosen in the pair CD. Thus, the best and worst choices provide information that can be “expanded” to several pairs of choices, with the higher ranked option being the one expected to be chosen. From a choice modeling standpoint, one can “stack” the pairs and code the option implied as the one chosen from each pair as a “1”, with the other option in the pair coded as a “0”. Standard conditional logit modeling estimation software can be used to estimate the model parameters. The results from our example are in Table 5.

The top part of Table 5 shows best and worst counts for each of the six destinations summed across all 420 respondents and 10 comparison sets, and the best minus worst counts (B − W) for each destination, as well as the self-reported choices (RP, or the average number of trips per person to each destination). The lower part of Table 5 contains the statistical results from estimating a conditional logit model from the best choices and the statistical results from estimating an OLS regression model (a linear probability model) from the best choices.

Fig. 1 plots the set of estimates from each of three regression models together with the revealed preference data against the best-minus-worst scores. The ordinary least squares regression model estimates (best-fit line), relating each of these four sets to the best-minus-worst scores are given below it. These estimates, together with associated R-squared values, suggest that all four sets are strongly linearly related to the best-minus-worst scores. In other words, all provide the same relative scale (measurement) position information about the objects. The takeaway from Fig. 1 is that one typically can estimate the latent scale positions (measures) of each object with any of the methods illustrated. Thus, one may wish to use the simplest approach by simply calculating best counts minus worst counts (BWS scores). One can calculate BWS scores for each person in a sample and describe the resulting distribution of the scores with typical statistics, such as the means, medians, standard errors, etc. (as shown in Table 5 for means and associated standard errors). It is worth noting that one is unlikely to need the standard errors of the BWS scores for each person because one rarely (if ever) needs to conduct statistical inference for one person. Instead, one wants to summarize the statistical properties of the sample. Fig. 1 also contains a graph of the relationship between the best-minus-worst (B − W) scores and the actual trips (RP). The relationship is approximately linear, with an R-squared of 0.86, suggesting that the BWS measures match well the reported choices.

If one wishes to conduct statistical inferences, one must be sure that the sample size is consistent with the desired test of parameter equality or differences. We discuss a general way to determine sample size requirements for best–worst studies later in the paper. The obvious caveat or limitation to the simple BWS score analysis is that averages can obscure underlying differences in measures across people. If one needs to understand individual differences, one can analyze the individual-level best–worst choices in several ways to gain

---

**Table 5**

Results for getaways study.

<table>
<thead>
<tr>
<th>Destinations</th>
<th>N</th>
<th>Calculated from best and worst choice totals</th>
<th>RP (ave reported trips)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Best choices</td>
<td>Worst choices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sum</td>
<td>Mean</td>
</tr>
<tr>
<td>Sydney</td>
<td>490</td>
<td>939</td>
<td>1.916</td>
</tr>
<tr>
<td>South Coast</td>
<td>490</td>
<td>768</td>
<td>1.567</td>
</tr>
<tr>
<td>Bowral</td>
<td>490</td>
<td>488</td>
<td>.936</td>
</tr>
<tr>
<td>Barrington</td>
<td>490</td>
<td>643</td>
<td>1.312</td>
</tr>
<tr>
<td>Katoomba</td>
<td>490</td>
<td>576</td>
<td>1.176</td>
</tr>
<tr>
<td>Central Coast</td>
<td>490</td>
<td>1347</td>
<td>2.749</td>
</tr>
</tbody>
</table>

---

such insights. That is, one can consider using various cluster analytic methods, latent class models or random parameter models. We illustrate the use of scale-adjusted latent class models (Magidson & Vermunt, 2007) later in the paper, but this is only one of several options open to analysts. The topic of capturing individual differences is vast, so in the interests of focus and space, we merely note that if one wants to explore individual differences, one must consider one or more methods for doing so. As a final comment, however, it is worth noting that for any research problem requiring reliable estimates of individual differences, one typically requires larger sample sizes than if one only wants to estimate sample average subjective values.

3.5. Examining choice frequencies across the sample

As explained above, BWS difference scores create a rank order of preference based on sample averages. We also fit individual-level LPM models to each person’s (respondent’s) BWS scores. The sample statistics for these individual-level LPM estimates are in Table 6. The results suggest that the Central Coast is the most preferred destination and that Bowral in the Southern Highlands of NSW is the least preferred. Other destinations are intermediate. The average of the respondents reported most recent visits (RP) are included in the table, and the correlation between mean and RP estimates is 0.91. Naturally, both the BWS scores and the RP reported visits contain errors; hence, the observed correlation should be considered merely as evidence that the relationship is relatively strong and in the correct direction. More generally, however, the standard errors of BWS scores can be calculated from the sample data to provide an indication of the degree of agreement in the sample on the subjective position of each object or item being measured. Lower standard errors imply greater agreement in the sample.

3.6. Sample sizes

BWS measures are derived from multinomial frequency counts or proportions. As noted earlier, we can calculate differences in frequency counts and ratios of frequency counts. We assume that one is not interested in inferences about any one individual associated with BWS scores, as one typically is not interested in inferences about one set of rating scale values from one person. Thus, sample sizes matter only to the extent that one designs a BWS task in such a way that one can calculate the scores of interest. In most cases, and at a minimum, BWS scores provide a full or partial ranking of items or objects of research interest.

Sample size considerations arise if one wants to make inferences about a population represented by a sample of people and/or if one wants to compare estimates (e.g., BWS scores) from two or more samples or different subsamples from a particular population (e.g., different “segments” in a sample). In these cases, BWS measures follow the rules for sample sizes associated with multinomial distributions. No specific methods for sample size have been developed for B–W scaling. However, sample size methods developed for multinomial proportions data can be used (e.g., Rose, 2011; Thompson, 1987). Several papers consider the issue of sample size requirements for multinomial proportions data (e.g., Angers, 1979, 1984, 1989), and Thompson (1987) derived a formula to calculate sample size requirements for multinomial proportions data. Thompson showed that, similar to the binomial case, sample size is a function of the level of acceptable error and the degree of desired confidence required by analysts in obtaining true population proportions. He also demonstrated that the sample size requirement for multinomial proportions data is independent of the number of multinomial categories (J outcomes or choice options; i.e., items, things) but not independent of what he termed the “worst possible outcome”. The worst possible outcome is defined as m of the J options having equal choice frequencies (proportions) or shares of 1/m, with the J − m remaining options having a value of zero.

Unfortunately, the value of m is not independent of the value of the confidence level, and therefore, it must be calculated for different levels of confidence required. For example, if we use Thompson’s approach and require a sample to satisfy the probability that at least 0.95 of all proportions are within 0.05 of the true population proportions and assume that m equals three, the required sample size is 510 respondents (independent of the number of options J). That is, in this example, the worst possible outcome that would be observed to occur is one where three population proportions are equal to 1/3, and the rest are equal to zero. Unfortunately, as things stand, we do not know whether different sample size implications are associated with different analysis methods. Thus, researchers may wish to rely on sample size estimation methods for discrete choice models, such as that from Rose (2011) for the general case of best only choices. It is worth noting, however, that it is likely that sample size requirements for estimating case 1 parameters using discrete choice models will be less than that for best only choices because BWS choices provide extra choice data for any given sample size.

4. Empirical example 2

We now turn our attention to a second empirical example that involves significantly large numbers of objects or items (72). The purpose of this example is to show that relatively large set sizes can be readily incorporated into a BWS study by reducing the size of the choice set through the use of a nested BIBD design. There are two reasons why researchers may wish to consider reducing comparison set sizes. 1) If individuals are relatively consistent in their choices, large numbers of objects per choice question give insufficient data points for middle-ranked objects (i.e., zero best-minus-worst counts will be observed, giving no information on relative preferences for middle-ranked objects, as was the case in Table 4). 2) Small choice sets make the task of choosing the best and worst easier for individuals who may have cognitive limitations.

There are two ways to collect enough choice data to derive an acceptable ranking when the choice set size is large. 1) After asking an individual for their best and worst objects in the set, one can ask a second round (or more) of best–worst questions to obtain additional ranking information (second best object, second worst, third best, third worst, etc.). Additional rounds of questions can be used to obtain a complete ranking of all objects in the choice set. This is particularly easy in web-based surveys because one can eliminate already chosen options from screens, making the task easier for respondents. If one uses this approach to collect additional best and worst choices, one must expand the choices to implied choice sets, and use more advanced analytical methods than the simple analytical methods discussed in this paper. That is, as noted by Horsky and Rao (1984), one can expand the data in each BIBD set to paired choices implied by the full or partial ranking obtained. As previously noted, the subjective values to be estimated are intercepts or alternative-specific constants in conditional logit models or more complex choice models, such as latent class or Generalized Multinomial Logit Models (or G–MNL, as per Fiebig, Keane, Louviere, & Wasil, 2010).

Whether and how many additional rounds of best and worst questions are needed will depend on a) the size of the comparison (choice) sets, b) how many objects (total choice options) are being measured and c) how critical it is for one to reliably and accurately measure (scale, estimate) each object of interest on the underlying latent scale.

<table>
<thead>
<tr>
<th>Destination</th>
<th>N</th>
<th>Mean</th>
<th>StdErr</th>
<th>StdDev</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney</td>
<td>420</td>
<td>0.734</td>
<td>0.030</td>
<td>0.624</td>
<td>5.138</td>
</tr>
<tr>
<td>South Coast</td>
<td>420</td>
<td>0.728</td>
<td>0.024</td>
<td>0.501</td>
<td>5.534</td>
</tr>
<tr>
<td>Bowral</td>
<td>420</td>
<td>0.575</td>
<td>0.022</td>
<td>0.453</td>
<td>1.186</td>
</tr>
<tr>
<td>Barrington</td>
<td>420</td>
<td>0.613</td>
<td>0.028</td>
<td>0.580</td>
<td>0.988</td>
</tr>
<tr>
<td>Katoomba</td>
<td>420</td>
<td>0.644</td>
<td>0.022</td>
<td>0.450</td>
<td>0.593</td>
</tr>
<tr>
<td>Central Coast</td>
<td>420</td>
<td>1.019</td>
<td>0.028</td>
<td>0.574</td>
<td>8.390</td>
</tr>
</tbody>
</table>

These results expand on the LPM ones of Table 5.
for each person. For example, if one only needs to accurately discriminate the best and worst objects and is less interested in intermediate objects (e.g., one often wants to sort potential product attributes for choice experiments into clearly important, intermediate and clearly unimportant), asking only one round of best and worst questions should be sufficient. More accuracy in differentiating and measuring the objects will dictate how many rounds of questions are needed, and this typically can be determined by a small pilot study in advance of the primary data collection.

A way to address large numbers of items to be measured is to use nested BIBD designs. That is, one first uses a suitable BIBD to generate a block of choice sets with a relatively large set size, and then uses a second suitable BIBD to reduce the number of objects in each choice set to a more manageable size. We illustrate the nested BIBD approach below and provide an example of how to implement it.

4.1. Issue of interest

There are several ways to ‘judge’ the quality of academic work. Stremersch, Verniers, and Verhoef (2007) examine the quality of individual articles, while Lehmann (2005) provides a range of criteria against which the quality of journals may be measured using some objective criteria such as citations and impact factors and other more subjective criteria. We approach the issue of measuring subjective evaluations of the quality of academic marketing journals, which forms the basis of our second example of best-worst scaling. We consulted several available journal ranking reports in marketing and business to obtain a fairly comprehensive set of academic and quasi-academic marketing journals (e.g., Fry, Walters, & Scheuermann, 1985; Geary, Marriott, & Rowlinson, 2004; Sivadas & Johnson, 2005; Starbuck, 2005). We focused on marketing-specific journals, excluding non-marketing journals such as Econometrica or Psychometrika, although marketing academics publish in them. This generated a list of 73 journals for BWS (see Appendix B).

4.2. Implementing BWS and design

To generate an appropriate BIBD, we consulted Street and Street (1987) to obtain a BIBD for 73 objects that yielded 72 comparison sets with 9 journals per set. We then used a second BIBD to expand each comparison set of 9 journals into 12 comparison sets with 3 journals per set. This reduced the size of comparison sets from 9 journals per set to 3 per set, yielding a total of $12 \times 72 = 864$ comparison sets for the entire survey.

Below is a simple example of how to pool two BIBDs in this way that involves a BIBD for 13 objects that produces 13 sets of size 9 and a BIBD for 9 objects that has 12 sets of size 3. We expand each block (row, set) in the first BIBD by using the second BIBD to make 12 new sets for each row in the first BIBD. This leads to $13 \times 12 = 156$ sets of size 3, as shown in Fig. 2.

Respondents were recruited by placing notices in online newsletters via postings to the ELMAR virtual community, the European Marketing Academy, The Australia–New Zealand Marketing Academy, and The Academy of Marketing in the United Kingdom. We also placed notices on the posting board at the Marketing Science Conference. In 2006, we published the survey online on a secure Web server and directed respondents to a URL with the survey. Questions were added to the survey to determine respondents’ academic rank, years in academia, geographic location and research specialties. To collect the BW choice data, we randomly assigned each survey respondent to one block in the first BIBD and asked the respondent to complete all 12 comparison sets from the second BIBD.

Approximately 900 respondents received an email directing them to the survey. A total of 529 respondents to the recruitment invitation and provided complete data (we deleted approximately 15 people whose surveys were incomplete or unusable, including one person who completed the survey 10 times).

4.3. BWS response data and simple analyses

As with example 1, the analysis of the journal ranking choices is based on assigning the best journal in the set a ‘+1’ and the worst a ‘−1’. The best and worst choice data are summarized by counting the occurrence of best and worst choices for each choice object (journal). However, as respondents were randomly assigned to each of the 72 initial blocks of 12 journals before being asked to evaluate all 9 blocks of three journals per initial block of 12, there were large differences in sample sizes for various survey versions. Thus, the resulting aggregate sample probability of choosing a journal (approximated by the proportion of choices for each journal) is independent of the probability that it is available to be chosen. To account for this, we reweighted the choice frequency counts for best and worst choices to take into account the probability of journal occurrence. The weights are calculated as follows:

$$\text{weight}_j = \frac{\text{average appearances for all journals}}{\text{average appearances for journal } j}.$$  

For example, suppose there are four blocks or versions of the best–worst task, which will occur if one randomly assigns sets of choice sets to the 4 versions in equal numbers without replacement. Each of the four versions constitutes a separate “survey”. Suppose one recruits participants from an online panel and randomly assigns each person who agrees to participate to one of the versions. Let the versions be V1, V2, V3 and V4. Let the number of respondents to each version be, respectively, 35, 20, 30 and 15, for a total of 100 participants. The versions need to be weighted because in the overall sample, some of the J objects being measured occur more/less often than others. As the average sample size = 25, the weights are calculated for each version as follows:

- Weight for V1 = $25/35 \approx 0.71$
- Weight for V2 = $25/20 = 1.25$
- Weight for V3 = $25/30 = 0.83$
- Weight for V4 = $25/15 = 1.67$.

One weights each observation in each version by the weights calculated above. We use this weighting approach in the example that follows.

The re-weighting increases the number of choices of journals appearing less often and decreases the number of choices of journals appearing more often. Using these weighted best and worst counts, we obtained a simple scale by subtracting weighted worst counts from weighted best counts. The aggregate sample results are shown in Appendix C, which also lists journals and occurrence-weighted BWS journal quality scores. The BWS scores for each journal in Appendix C gives a ratio scale of journal quality, meaning that one can conclude that a journal with a score of 1.00 has approximately twice the level of perceived quality of a journal with a score of 0.50. Because scores are ratio-scaled, they can be transformed into point systems consistent with measured levels of perceived quality.

4.4. Further examination of the sample

Our sample contained enough choice data to also conduct separate analyses for each of three regions, North America (Canada and the USA = 220), Europe (94) and Australia–New Zealand (107). We tested the hypothesis that there is heterogeneity in how academics perceive (value) journal quality. To do this, we estimated a scale-adjusted latent class model (SALCM) from the choice data (Magidson & Vermunt, 2007). The results of the SALCM produced only one journal-quality class using the Bayesian information criteria for model selection. That is, respondents tended to rank journals similarly in all three regions. Discussion and examples of the SALCM estimation technique can be found in Burke et al. (2010) and Flynn, Louviere, Peters, and Coast (2010). Despite our finding that respondents ranked journals similarly across regions, it is possible that researchers in top schools would rank journals differently from researchers in other schools. School level data were...
not available in this study, but examining such heterogeneity across schools may provide an interesting line of inquiry for future researchers.

The top 10 journals are shown in Table 7. The sample mean quality measure for all journals is 41.5 (see Appendix C), with a standard error of 2.73; hence, differences of 5.5 scale units are significant. Accordingly, the Journal of Marketing and Journal of Marketing Research are perceived as top journals and statistically equal firsts, followed by the Journal of Consumer Research, Journal of the Academy of Marketing Science and Marketing Science, which are statistically equal seconds in perceived quality. The next four are statistically equal thirds, with the Journal of Advertising significantly lower than the top two in this tier.

5. Discussion and conclusions

A Google search for the terms “best–worst scaling” and “maximum difference scaling” returned, respectively, 1.96 M and 3.7 M “hits” (April 22, 2012). Perusal of the first several pages of hits clearly shows that the choice-based measurement approach that we call “best–worst scaling” is used primarily by academics, while practitioners are mainly included in the hits for “maximum difference scaling”. The results also suggest that a growing number of academics in several different fields are adopting the BWS approach. Apart from the original Finn and Louviere (1992) paper that introduced the approach and the theoretical treatment by Marley and Louviere (2005) that derived the formal measurement properties of various case 1 best–worst choice models, there are few “how to do it” papers. Thus, the objective of this paper was to describe and discuss ways to implement, analyze and interpret case 1 (object case) best–worst scaling (BWS) applications, and show how to use BWS by applying it to two empirical examples.

We showed that BWS is relatively easy to implement and analyze, even with fairly large numbers of objects (e.g., the academic journals example), making it accessible to many academic and applied researchers. We focused on simple ways to analyze BWS data (using best-minus-worst scores and simple regression models) to show that one can obtain good results with fairly simple analysis methods. A more detailed and formal treatment of the theory underlying BWS is in Marley and Louviere (2005); our aim was to give as straight forward an explanation of the theory and methods as possible. Thus, we emphasized simple counts of choices (i.e., sums), expansion of best and worst choices to implied paired choices and graphical tests to allow one to assess if the data are consistent with theory.

Table 8 below illustrates the steps in designing a best–worst scaling study.

5.1. BW scaling for groups of individuals

We also applied more complex regression models frequently used to analyze data from traditional discrete choice experiments. The measurement values estimated by these conditional logit models are the natural logarithms of the classical (Luce, 1959) multiple-choice model that yields ratio scales. We caution that many of the same issues associated with those models also apply to BWS. Most notably, the assumptions underlying such models are quite strong and have been discussed in the discrete choice modeling literature for many years (e.g., Train, 2003). Perhaps most importantly, such models theoretically apply only to single people; additional assumptions are required to extend them to aggregates of people. How well such models approximate individuals compared with aggregates of individuals remains unresolved. Recent work on choice
models for single individuals by Louviere et al. (2008) suggests that individual-level models can outperform more aggregate models. Thus, it may be that McFadden's (1974) conditional logit model or Luce's (1959) model may be a reasonable first approximation to a person's unknown choice processes for BWS choice tasks. Further work is needed to understand if and when one needs to, and if so, how to, relax assumptions associated with these models. We set the latter issues aside here and merely note that a great deal of experience with BWS over the past five years suggests that these models seem to be reasonable first approximations for unknown individual-level choice processes.

5.2. Unresolved issues with BWS

There are a number of unresolved issues with BWS that can be viewed as future research opportunities. For example, because BWS relies on discrete choices, it has the limitations of random utility choice models, such as possible violations of the independence of irrelevant alternatives (IIA) property of Luce's (1959) model and McFadden's (1974) conditional logit model. Whether IIA violations exist is an empirical issue, and how serious they are in individual-level BWS choice data remains an open issue. Nevertheless, we note that the equal co-occurrence property inherent to BIBDs allows one to estimate violations of the IIA property of simple choice models as it ensures that the two-way interactions (cross-effects) are estimable (Lazari & Anderson, 1994).

Similarly, objects or items of interest in BWS applications may exhibit various degrees of similarity and/or correlated errors, which is also an open issue. Those familiar with discrete choice models will recognize that these issues have been widely discussed in the choice-modeling literature. Hence, the issues are easy to state but complex to resolve, especially when one is modeling single people.

We also should emphasize the need to consider decision rules used by respondents. Because individuals are asked to choose the best and the worst objects (i.e., largest/smallest, most/least preferred, etc.), BWS is sometimes called 'maximum difference' scaling (or 'maxdiff') scaling. The latter nomenclature is unfortunate because the maxdiff model is only one of a number of models in the process that an individual might use to make a series of best—worst choices; in mathematical psychology, the maxdiff model assumes that an individual considers all possible best—worst pairs (simultaneously) and chooses that pair that maximizes the difference between the two objects comprising the pair. Naturally, an individual might use alternative choice processes. For example, they might choose the best stimulus first, followed by the choice of the worst stimulus; or they might choose the worst stimulus first, followed by the best stimulus, and so forth. The way(s) individuals make such choices is an empirical question. However, each way implies a different possible process model of their choices and, strictly speaking, a different statistical model.

We note that current BWS scales are relative to the sets of objects studied. For example, if we offer a person the set (Hitler, Mussolini, and Stalin) and ask the person to choose the best and worst national leaders, they would make two choices. However, it is likely that, if asked, they would say that no one in the set was a 'good' leader. Ongoing work aims to resolve this problem by developing BWS measures that reflect absolute as well as relative positions. More generally, however, solutions to this problem require extra information external to a BWS task, and theoretically sound solutions to this problem would be welcome. Meanwhile, one can ask people to report whether “none of the objects is good” and/or “none of the objects is bad”, which can be used to anchor the scales. The latter is common practice in discrete choice experiments dating from Louviere and Woodworth (1983). Another possibility is anchoring relative to a status quo option in each comparison set, but to our knowledge this has, as yet, not been conducted.

We illustrated and discussed simple and more complex ways to analyze best—worst choice data. Examples evaluating a small and large number of objects were demonstrated, thus illustrating the generalizability of the methods. Future work should examine the extent to which BWS can out-perform traditional rating scales and investigate whether the benefits noted herein are generalizable across policy and marketing areas.

Acknowledgments

We gratefully acknowledge the assistance of Tony Marley and Andrew Kyngdon for the collection of the data on weekend getaways used in this paper.

This paper was prepared with the assistance of the Services Innovation Research Program, QUT Business School, Queensland University of Technology.
Appendix A. Histograms for individual-level LPM estimates

![Histogram for Sydney](image1)

![Histogram for SoCost](image2)

![Histogram for Bowral](image3)

![Histogram for Katoomba](image4)

![Histogram for CenCoast](image5)

![Histogram for Barrington](image6)
Appendix B. Journals — in alphabetical order

<table>
<thead>
<tr>
<th>Academic Marketing Journals—overall scores and regional scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall sample</strong></td>
</tr>
<tr>
<td>100.00</td>
</tr>
<tr>
<td>98.01</td>
</tr>
<tr>
<td>91.19</td>
</tr>
<tr>
<td>90.06</td>
</tr>
<tr>
<td>89.47</td>
</tr>
<tr>
<td>80.30</td>
</tr>
<tr>
<td>78.69</td>
</tr>
<tr>
<td>75.27</td>
</tr>
<tr>
<td>74.72</td>
</tr>
<tr>
<td>71.74</td>
</tr>
<tr>
<td>68.47</td>
</tr>
<tr>
<td>67.74</td>
</tr>
<tr>
<td>66.08</td>
</tr>
<tr>
<td>65.67</td>
</tr>
<tr>
<td>63.65</td>
</tr>
<tr>
<td>59.30</td>
</tr>
<tr>
<td>57.17</td>
</tr>
</tbody>
</table>

Appendix C (continued)

<table>
<thead>
<tr>
<th>Overall sample</th>
<th>North America</th>
<th>Aust/NZ</th>
<th>Europe</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.96</td>
<td>41.26</td>
<td>60.26</td>
<td>61.51</td>
<td>International Journal of Market Research</td>
</tr>
<tr>
<td>53.43</td>
<td>49.15</td>
<td>51.04</td>
<td>63.88</td>
<td>Journal of International Marketing</td>
</tr>
<tr>
<td>52.19</td>
<td>49.99</td>
<td>41.44</td>
<td>58.68</td>
<td>Industrial Marketing Mgmt</td>
</tr>
<tr>
<td>51.14</td>
<td>40.11</td>
<td>67.00</td>
<td>45.48</td>
<td>Academy of Marketing Science Review</td>
</tr>
<tr>
<td>49.59</td>
<td>59.29</td>
<td>30.53</td>
<td>22.64</td>
<td>Journal of Personal Selling and Sales Mgmt</td>
</tr>
<tr>
<td>50.21</td>
<td>54.48</td>
<td>52.77</td>
<td>25.44</td>
<td>Journal of Services Marketing</td>
</tr>
<tr>
<td>48.42</td>
<td>47.14</td>
<td>57.93</td>
<td>45.71</td>
<td>Journal of Product and Brand Mgmt</td>
</tr>
<tr>
<td>48.00</td>
<td>41.63</td>
<td>65.88</td>
<td>42.41</td>
<td>Journal of Consumer Behavior</td>
</tr>
<tr>
<td>46.71</td>
<td>53.23</td>
<td>24.98</td>
<td>39.92</td>
<td>Journal of Macromarketing</td>
</tr>
<tr>
<td>46.15</td>
<td>40.52</td>
<td>41.70</td>
<td>56.25</td>
<td>Marketing Theory</td>
</tr>
<tr>
<td>44.35</td>
<td>37.36</td>
<td>50.05</td>
<td>46.83</td>
<td>International Marketing Review</td>
</tr>
<tr>
<td>43.73</td>
<td>46.39</td>
<td>38.74</td>
<td>35.05</td>
<td>Journal of Marketing Theory and Practice</td>
</tr>
<tr>
<td>43.21</td>
<td>36.73</td>
<td>44.71</td>
<td>46.48</td>
<td>Journal of Marketing Mgmt</td>
</tr>
<tr>
<td>42.14</td>
<td>35.63</td>
<td>44.88</td>
<td>29.76</td>
<td>Journal of Strategic Marketing</td>
</tr>
<tr>
<td>39.38</td>
<td>40.65</td>
<td>41.29</td>
<td>29.62</td>
<td>Journal of Business-to-Business Marketing</td>
</tr>
<tr>
<td>38.52</td>
<td>33.14</td>
<td>41.90</td>
<td>42.76</td>
<td>Journal of Retailing and Consumer Services</td>
</tr>
<tr>
<td>37.95</td>
<td>44.35</td>
<td>29.87</td>
<td>23.74</td>
<td>Journal of Consumer Marketing</td>
</tr>
<tr>
<td>37.12</td>
<td>28.56</td>
<td>35.78</td>
<td>49.82</td>
<td>Journal of Business and Industrial Marketing</td>
</tr>
<tr>
<td>36.86</td>
<td>28.52</td>
<td>44.89</td>
<td>54.99</td>
<td>International Jl Retail &amp; Distribution Mgmt</td>
</tr>
<tr>
<td>36.34</td>
<td>22.97</td>
<td>38.58</td>
<td>41.05</td>
<td>Journal of International Marketing and Marketing Research</td>
</tr>
<tr>
<td>36.24</td>
<td>28.89</td>
<td>22.44</td>
<td>29.51</td>
<td>Journal of Interactive Marketing</td>
</tr>
<tr>
<td>36.26</td>
<td>32.81</td>
<td>34.35</td>
<td>24.43</td>
<td>Journal of Global Marketing</td>
</tr>
<tr>
<td>31.98</td>
<td>20.80</td>
<td>37.74</td>
<td>35.92</td>
<td>Marketing Intelligence and Planning</td>
</tr>
<tr>
<td>29.13</td>
<td>19.91</td>
<td>26.46</td>
<td>18.85</td>
<td>Journal of Consumer Behavior</td>
</tr>
<tr>
<td>29.11</td>
<td>31.86</td>
<td>28.44</td>
<td>9.60</td>
<td>Services Marketing Quarterly (aka Jl Prof'l Services Mktg)</td>
</tr>
<tr>
<td>28.03</td>
<td>28.14</td>
<td>23.35</td>
<td>18.79</td>
<td>Jl of Targeting, Measurement &amp; Analysis for Mktg</td>
</tr>
<tr>
<td>27.25</td>
<td>22.52</td>
<td>38.16</td>
<td>13.79</td>
<td>Qualitative Market Research: An International Journal</td>
</tr>
<tr>
<td>26.96</td>
<td>26.08</td>
<td>21.69</td>
<td>20.03</td>
<td>Advances in International Marketing</td>
</tr>
<tr>
<td>26.67</td>
<td>27.13</td>
<td>23.66</td>
<td>19.92</td>
<td>Journal of Marketing Channels</td>
</tr>
<tr>
<td>26.50</td>
<td>20.57</td>
<td>28.28</td>
<td>17.26</td>
<td>Journal of International Consumer Marketing</td>
</tr>
<tr>
<td>26.48</td>
<td>22.28</td>
<td>34.45</td>
<td>13.89</td>
<td>Journal of Marketing Communications</td>
</tr>
<tr>
<td>24.46</td>
<td>13.74</td>
<td>40.89</td>
<td>16.57</td>
<td>Australasian Marketing Journal</td>
</tr>
<tr>
<td>23.95</td>
<td>21.75</td>
<td>26.86</td>
<td>12.38</td>
<td>Journal of Non Profit and Public Sector Marketing</td>
</tr>
<tr>
<td>23.21</td>
<td>21.56</td>
<td>22.56</td>
<td>20.65</td>
<td>Academy of Marketing Studies Journal</td>
</tr>
<tr>
<td>22.07</td>
<td>19.61</td>
<td>35.72</td>
<td>10.09</td>
<td>Australian Journal of Market Research</td>
</tr>
<tr>
<td>21.55</td>
<td>24.97</td>
<td>15.06</td>
<td>10.43</td>
<td>Marketing Education Review</td>
</tr>
<tr>
<td>20.77</td>
<td>20.43</td>
<td>14.14</td>
<td>18.88</td>
<td>Database Marketing Journal</td>
</tr>
<tr>
<td>18.89</td>
<td>9.61</td>
<td>34.11</td>
<td>10.74</td>
<td>Asia-Pacific Journal of Marketing and Logistics</td>
</tr>
<tr>
<td>17.81</td>
<td>20.71</td>
<td>13.94</td>
<td>16.77</td>
<td>International Jl of Nonprofit &amp; Voluntary Sector Mktg</td>
</tr>
<tr>
<td>16.91</td>
<td>16.60</td>
<td>16.27</td>
<td>10.23</td>
<td>Marketing Health Services (aka Jl Health Care Mktg)</td>
</tr>
<tr>
<td>16.31</td>
<td>14.52</td>
<td>11.68</td>
<td>14.95</td>
<td>Journal of Marketing for Higher Education</td>
</tr>
<tr>
<td>16.01</td>
<td>19.73</td>
<td>10.53</td>
<td>6.05</td>
<td>J of Euromarketing</td>
</tr>
<tr>
<td>12.35</td>
<td>6.95</td>
<td>15.77</td>
<td>4.75</td>
<td>Marketing Bulletin</td>
</tr>
<tr>
<td>10.10</td>
<td>8.26</td>
<td>5.48</td>
<td>17.03</td>
<td>International Journal of Bank Marketing</td>
</tr>
<tr>
<td>9.59</td>
<td>11.83</td>
<td>3.87</td>
<td>3.12</td>
<td>Sport Marketing Quarterly</td>
</tr>
<tr>
<td>7.73</td>
<td>1.60</td>
<td>6.37</td>
<td>3.11</td>
<td>Journal of Fashion Marketing and Mgmt</td>
</tr>
<tr>
<td>7.32</td>
<td>4.90</td>
<td>2.22</td>
<td>0.00</td>
<td>Journal of Vacation Marketing</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>8.11</td>
<td>Marketing Week</td>
</tr>
</tbody>
</table>

References


Consumer research suffers from a lack of respect for data. Researchers routinely fail to report full experiments that do not produce expected results and often eliminate alleged “outliers” on the basis of inappropriate rules, leading to biased test reports. Scholars appear to be relying less on non-experimental data, even as the serious limitations of experimental data may create structural discrepancies with the other, non-experimental cases of a phenomenon or process, such that it becomes impossible to study some major consumer phenomena. The lessons from empirical data get accepted only when they can be described as confirming preexisting conceptual frameworks. This article presents an extensive analysis of multiple forms of a lack of respect for the data and proposes some remedies. Overall, data should never play a subordinate part.

1. Introduction

Two recent, widely reported scandals involved complete forgeries of data sets by consumer researchers: the Stapel case (investigated in depth by the Levelt Committee, Noort Committee, & Drenh Committee, 2012) and the Smeesters case (Retractionwatch, 2012; Simonsohn, 2011). But the data-related lessons to be derived from these extreme cases should not be limited just to avoiding obvious forgeries. Rather, a lack of respect for data takes various, often more subtle forms but can be just as dangerous. With this article, I attempt to draw the attention of my consumer behavior colleagues and doctoral students to the threat that insufficient respect for data creates for consumer behavior research.

In a citation that rightfully caused great consternation for the Levelt Committee et al. (2012), Stapel (2000, p.6) claimed: “The leeway, the freedom we have in the design of our experiments is so enormous that when an experiment does not give us what we are looking for, we blame the experiment, not our theory. [At least, that is the way I work.] Is this problematic? No.” This citation clearly expresses Stapel’s belief in the “primacy of the theory—and therefore the subordinate role of the data?” (Levelt Committee et al., 2012, p. 40). But data are not subordinate. In this introduction, I illustrate the dangers of treating data as subordinate with a few examples. I reserve the in-depth discussions and recommendations for subsequent sections.

1.1. Do not mutilate the data set

We must beware unwarranted modifications of a data set after it has been collected. Beyond the (hopefully) rare forgery of observations, I am concerned about data elimination practices that discard certain observations or even full data sets. Consider the “best of” tactic: a common research approach is to run multiple experiments but report only some of them, namely, those that support the researcher’s argument, or else collect multiple measures of a dependent variable and then report only a selected subset of the results. This selective reporting tactic leads to dangerous biases in reported hypothesis tests. Another dangerous behavior eliminates alleged outliers, before performing analyses of variance (ANOVA) on experimental data, using inappropriate rules that likely produce an upward bias in the F-tests without ensuring that the reduced data set follows a Gaussian distribution. Both “best of” tactics and unwarranted discarding of alleged outliers produce articles that analyze and report mutilated, rather than the original, data sets.

1.2. Avoid discrepancies between experimental data sets and other instances of the phenomenon or process under investigation

The phenomena and underlying processes we research existed prior to our investigations, occur in many other settings than the ones in which we investigate them, and will continue long after we are done with our project. These are truisms that researchers must keep in mind. Considering the dominance of experimental research in consumer research, we need to be concerned about possible discrepancies between the data in our experiments and data relative to non-experimental instances of the phenomena or processes being investigated. Discrepancies might arise because consumer behavior experiments tend to be myopic in a temporal sense, mostly analyzing short-run answers to short-run stimuli, whereas consumers typically develop knowledge and choice strategies over time and multiple, repeated choice occasions. Other discrepancies may occur when the study results come from a narrow
laboratory setting or a very specific student population, without any confirmation by replications in other settings, whether undertaken by the author or encouraged by reviewers. Some central aspects of consumers' behaviors cannot be studied in a lab. Consider, for example, the changes that have occurred as Chinese consumers increasingly have moved to large cities and found ways to earn higher incomes.

1.3. Let the data speak outside of their predefined script

If we listen to data only when they speak consistently with a predefined schema or confirm hypotheses that are based on previous research or an a priori conceptual framework, researchers encounter another danger. Instead, we need to listen to the data even (or particularly) when they offer unexpected lessons, especially if we have no solid preexisting theory (which is not infrequent in consumer behavior or, more broadly, in marketing). In such cases, we must remain open to learning new lessons and maintain an abductive approach. For example, if we observe a lack of convergence across different measures of a theoretically unidimensional construct, we may reconsider this unidimensionality rather than eliminate divergent measures. Similarly, we should be cautious about possible nonlinearities and collect data in a manner that reveals them.

1.4. Finding ways to address these threats

The problems I highlight in this article sometimes are driven by specific practices or requests made by review teams; I therefore discuss this specific aspect when appropriate. Of course, it also would be nice to justify my assertions by providing empirical evidence of these wrongdoings. But the guilty seldom report their crimes. Authors rarely mention data in a manner that reveals them. We should be cautious about possible nonlinearities and collect more instances. But the guilty seldom report their crimes. Authors rarely mention data in a manner that reveals them.

2. Mutilated data sets

Three forms of lack of respect for the data occur at the analysis stage: hidden or unwarranted elimination of observations, selective reporting of experiments, and “best of” tactics. These concerns suggest the need for better reciprocal controls by coauthors.

2.1. Concealed or unwarranted elimination of observations

It is frequently necessary to “clean” a data set and eliminate certain observations before performing the statistical analysis. However, this process should be transparent and undertaken for appropriate reasons. To provide empirical support for the claims in this section, I reviewed all 72 experimental articles in the most recent volume of Journal of Consumer Research (JCR Vol. 39, June 2012–April 2013). I counted 53 instances in which the authors reported eliminating observations for a specific reason (for details, see the Web Appendix). The allocation of these instances across articles reveals a troubling pattern: the average frequency is 53/72 = .74 instances per article, so if they follow a Poisson process, we should expect to observe 34.49 articles with no instance, 25.39 with a single instance, and 12.12 articles with two or more instances. But the pattern in reality is very different, featuring too many papers with no instances (51), too few papers with a single instance (4), and too many with two instances or more (17). These differences are significant ($\chi^2 = 27.89$, 2 df, $p = 4.4 \times 10^{-7}$).

Why do so many articles report no data elimination at all (51 out of 72, or 71%)? Diverse factors might lead to the elimination of observations, often corresponding to unforeseen, incidental deviations from the scheduled data collection protocol (e.g., computer problem, failure to follow instructions, refusal to perform experimental tasks, missing answers, sickness; see the Web Appendix), so it seems unlikely that all the observations collected for all 51 articles were immediately perfect—especially considering that they generally include at least four experiments each. Rather, at least some of these authors likely eliminated some original observations but did not report how or why. This conclusion is reinforced by the realization that, among articles that report at least one instance of eliminating an observation, more than 80% (17 of 21) reported two instances or more (and up to five in two articles). That is, when authors correctly report that they eliminated observations, they usually find more than one instance to report. Therefore, I conclude that the elimination of data is very common but mostly hidden. This conclusion is reinforced by the contrast between my review of articles that did not report any data elimination and my multiple informal discussions with colleagues who indicated that they considered preliminary data cleaning both routine and necessary (e.g., to detect respondents who did not read the instructions carefully). Thus, the first recommendation is to respect the data by always reporting clearly whether observations have been eliminated and, if so, how many and for what specific reason.

Eliminating an observation is legitimate if it is based on a documented incident in the data collection, as listed previously. But observations also might be eliminated solely because they appear to be “outliers,” in that they show extreme values on the dependent variable. These eliminations reportedly seek to bring the updated data set (i.e., after elimination) close to a normal (Gaussian) distribution and enable ANOVA. Two common rules guide these eliminations: eliminate observations that are more than 3 standard deviations away from (or above) the sample mean of the original distribution and eliminate observations that are more than 1.5 interquartile ranges (IQR) above the upper quartile of the original distribution or more than 1.5 IQR below its lower quartile. Before discussing these two rules, let me stress that nonparametric statistics would offer a robust, easy-to-use, alternative approach for data analysis that does not require eliminating observations with extreme values.

The rule that suggests eliminating observations that are more than 1.5 IQR beyond the quartiles relies on a statistical tool introduced by Tukey (1977): the “box-and-whiskers” plot, an example of which appears in Fig. 1. Panel A. To visualize the distribution of a numerical variable, Tukey recommends computing its median and quartiles to plot a “box” that contains the middle 50% of the distribution (top quartile to bottom quartile, with a horizontal bar indicating the median). Then Tukey defines two “whiskers” or “inner fences” at 1.5 IQR beyond the top and bottom quartiles, marked in the plot by two horizontal lines above and below the box, as well as two “outer fences” located a further 1.5 IQR beyond the whiskers (not represented on the plot). Observations between the whiskers and the outer fences are “outside,” while observations beyond the outer fences are “far out.” Recent applications change these names slightly, such that in SPSS, the box-and-whiskers plot is called the “boxplot,” outside observations are “outliers,” and far out observations are called “extreme cases.”

Tukey’s (1977, chapter 2) original text makes clear that he designed this innovative plot to provide an intuitive visualization of the distribution of a variable. For example, in Fig. 1, it is immediately obvious that the distribution in Panel A (of the performance of 167 respondents on a classical “speed of treatment” test) is symmetrical, close to a Gaussian, whereas the distribution in Panel B (of the quantities of an industrial product purchased by different customers) is highly skewed. But at no

---

1 There is an exception: if there is no observation beyond one of the whiskers defined as above, it moves closer to the median, at the level of the extreme observation.
point does Tukey suggest that observations located outside or far out should be eliminated. In multiple informal discussions, I have heard colleagues refer to Tukey to justify eliminating observations that they considered outliers, solely because they are located beyond the whiskers on the dependent variable.

By definition, an outlier is "a statistical observation that is markedly different in value from the others of the sample" (Merriam-Webster dictionary online). In a boxplot, an observation located beyond the whiskers also gets referred to as an outlier. But this similarity in terms does not imply that such a particular observation does not belong in the data set or should be discarded, and two primary reasons argue against such actions: this procedure strongly biases ANOVA results, yet it does not bring the updated data set (after elimination) into a normal (Gaussian) distribution.

Consider the two possible horns of this alternative: when the raw data set is Gaussian and when it is not.

First, if the population of interest is perfectly Gaussian, about 4.3% of the observations should be beyond the whiskers, and there is no reason to eliminate all of them automatically. A more surprising observation would be if there were no outliers. Eliminating such observations implies that the updated distribution is no longer Gaussian, because it will have lost its two tails. The estimator of the within-group variance, an essential component of the ANOVA procedure, thus gets biased downward by more than 20%. Because the resulting estimate is the denominator of the F-statistics, the suppression of the two tails would lead to overestimates of F by about 25%, changing marginal tests from non-significant to significant.

Second, if the distribution is not Gaussian, it might be lognormal or exponential, as is frequent in real-world populations. Various examples of naturally occurring lognormal distributions come from across the sciences (Atchinson & Brown, 1963; Johnson, Kotz, & Balakrishnan, 1994; Limpert, Stahel, & Abbt, 2001). Firm sizes follow lognormal distributions, for example, so consider the distribution in Fig. 1, Panel B (quantities purchased by different customers in an industrial market) to illustrate the absurd consequences of removing observations beyond the whiskers when the raw distribution is lognormal and skewed to the right. The distribution in Panel A has a measured skewness of .19 (SD = .19), consistent with the zero skewness of a theoretical Gaussian distribution; the distribution in Panel B reveals a measured skewness of 2.66 (SD = .21), which is significantly above zero. Eliminating all the observations beyond the whiskers in the data from Panel B produces a

---

**Fig. 1.** Examples of “Box-and-Whisker” Plots. A. Performance on a psychological test: single outlier. B. Quantities purchased: many high-value outliers. C. Quantities purchased after application of the alleged Tukey rule: still many high-value outliers. D. Natural logarithm of quantities purchased: no high-value outliers.
distribution of the remaining observations (Panel C) that is clearly not Gaussian and still inappropriate for ANOVA (skewness = 1.68, SD = .22). And this would underestimate the within-group variance, with a decrease in estimated variance from 15.2 million in the original data to 2.8 million, or more than 80%, prompting a massive upward bias in the estimated F statistics. Thus, in both Gaussian and non-Gaussian cases, removing observations located beyond the whiskers leads to underestimates of within-group variance and overestimates of F statistics in ANOVA.5

Another reason not to remove observations located beyond the whiskers is that it leads to absurdities for many distributions. Consider again the lognormal distribution in Fig. 1, Panel B. The upper whisker is located at 7382, so all major purchasers in this market might be eliminated systematically. Even more absurdly, the computed value of the lower whisker is negative, so a researcher applying this rule would eliminate no small purchasers. Together, the procedure creates a major bias: The mean drops by 42%, from 2479 to 1446.

Finally, this procedure should be rejected on the basis of an authoritative argument: Tukey (1977), who developed the definition of outliers on the basis of boxplots, never recommended excluding them. John Tukey is not Dr. Guiloin. Instead, he recommended a totally different approach for the “re-expression” of the original variable, which I will describe later.

Moving on to the next rule, which recommends eliminating observations that are more than 3 standard deviations above the sample mean of the original distribution, the Gaussian distributions are better, because this procedure removes only the top .13% of the distribution and just slightly biases the sample. But it still can lead to absurd consequences for other distributions. In a lognormal distribution similar to the one in Panel B, the rule eliminates the top 4% of the sample and still heavily biases the mean, which drops by 22% from 2479 to 1934. Moreover, the distribution remains non-Gaussian and heavily skewed (2.42, SD = .21), and the estimated variance gets biased downward by 52% (from 15.2 to 7.3 million). The F statistics obtained from this updated sample thus continue to be erroneously overestimated. In addition, the 3 SD rule suffers from circular reasoning, in that the mean and standard deviation used to decide which observations to eliminate get estimated from the original sample, including the extreme observations that they serve to eliminate subsequently.

In informal discussions, I have heard colleagues argue that alleged outliers should be identified and eliminated separately for each experimental condition, a position I consider logically inconsistent. Researchers running an experiment hope to test, using ANOVA, the null hypothesis that the distribution of the dependent variable is the same for all conditions. To treat each condition separately in a preliminary step requires the assumption that the conditions lead to different distributions of the dependent variable—that is, a rejection of the test of H0 that will be performed in the second step. It is logically inconsistent to test a hypothesis using a data set that has been modified already (in the elimination step) by the assumption that this self-same hypothesis is rejected.

In some cases, a review team suggests applying the same procedure regarding outliers to all studies in an article; it happened to me personally recently. I consider this recommendation inappropriate. Different studies in the same article may have unique dependent variables, such as the average of Likert scales (likely a Gaussian distribution), the amount consumers would be willing to pay (likely a lognormal distribution), and a duration (likely an exponential distribution). The procedure to handle alleged outliers cannot be the same in these distinct cases.

Finally, the removal of outliers offers ample opportunities to create different variants of a data set: the original, the set obtained after removing outliers identified over the full sample, or the set obtained after removing outliers identified within each experimental condition. Researchers could perform ANOVAs on all three sets and simply report the results that fit best with the conclusions they want to reach—a typical example of the “best of” tactics denounced by the Levelt Committee et al. (2012) and that I discuss further in Section 2.3.

In this context, respect for the data suggests several recommendations. First, researchers can eliminate observations if they can justify doing so by specific incidents that affected the data collection. They may not remove observations solely because they indicate extreme values, especially before the researcher assesses the overall shape of the raw distribution, which should always be the first step. If the distribution is close to Gaussian, the raw data should be analyzed without eliminating observations beyond the whiskers (though eliminating observations at more than 3 SD from the mean or beyond the outer fences might be reasonable). If the data are far from Gaussian, we should consider nonparametric tests as an alternative approach (for a good example, see Frederick, 2012, p.2). Another option would be to follow Tukey’s (1977) recommendation and “re-express” the original variable by transforming it, such that the distribution of the transformed variable is as symmetrical and close to Gaussian as possible. Tukey devotes a full chapter (Chapter 3) to describing possible transformations: logarithms, power functions, reciprocals, and so forth. For a good example, see Frederick (2012 p.17). In Fig. 1, Panel D, I provide another example, namely, the boxplot of the logarithm of the variable described in raw form in Panel B. These raw data contained many high value outliers in Tukey’s sense: 8.7% beyond the whiskers, including 5.8% beyond the outer fences. After re-expression, the distribution emerged as symmetrical and appropriate for an ANOVA. The percentage of outliers dropped to 1.4%, all located at the bottom end of the distribution, because no more high-value “outliers” existed. Company sizes follow lognormal distributions, so Panel D clearly shows that the major purchasers on this market belong in the distribution and in the analysis.

In practice, whenever researchers eliminate observations, they should include an appendix that describes precisely their argument for the elimination, as well as the full distribution of observations before and after elimination (i.e., values taken by the discarded observations). We should avoid succinct, non-informative statements such as, “For the pencils condition, two outliers were excluded” (Frederick, 2012, p. 5).

2.2. Hidden experiments

Discussions with colleagues suggest that it has become common, when preparing an article, to run many more experiments than what appears in the final article. In her ACR Presidential address, Kahn (2006) relied on an estimation that indicated that an A-level paper requires running three times as many experiments as what will appear in the published version. Simmons, Nelson, and Simonsohn (2011) describe how Francis (2012) argued that results presented by Galak and Meyvis (2011), in which seven of eight studies were positive, seemed “suspicious,” because the size of the effect indicated there should have been more negative studies. The latter authors replied that they actually had obtained five additional negative studies but kept them in their “file drawer” rather than reporting them; thus their results were no longer suspicious, taking all 13 studies into account. It appears that many authors believe that journal review teams demand uniformly perfect results—a belief that discourages authors from reporting non-significant or even marginally significant studies. This concurs with Stapel’s statement “that journal editors preferred simplicity. They are actually telling you: ‘Leave out this stuff. Make it simpler’” (Bhattacharjee, 2013, p. 4).

Respect for the data instead requires mentioning experiments that did not produce the expected results, rather than discarding them. They provide rich information, namely, empirical evidence designed and collected under the full control of the experimenter but contradictory with the predictions. It is the opposite of serendipity. If an effect...

5 Space considerations prevent a similar discussion for other non-Gaussian distributions. However, Sim, Gan, and Chang (2005, p. 650) demonstrate that, for detecting outliers, the usual values of 1.5 or 3 IQR are “completely inappropriate for a skewed distribution, such as the exponential distribution.”
is so weak that it is significant in only four experiments out of eight, this is informative and should be reported. If the effect appears only with certain manipulations, measures, populations, experimental settings, and so forth, this too is informative and should be reported. Thus, an abductive approach may be the key. Researchers should develop hypotheses about why carefully designed, well-controlled experiments fail to produce the expected results, then test them. Reporting all experiments would help avoid the situation described earlier by Simmons et al. (2011). Galak and Meyvis (2012) asserted that Francis (2012) should have asked them about their “file drawer” studies instead of writing an article criticizing them. Ultimately, it would have been simpler and more straightforward for them to have reported all 13 studies in the initial article.

This problem also might reflect the blurred distinction between pilot tests and full-scale experiments, due to the wide acceptance\(^6\) (or even recommendation!) of the use of student samples or Mechanical Turk (MTurk)\(^7\) samples that lower the cost of experiments.\(^8\) As a consequence, researchers can quickly run an experiment, examine the results, and decide later whether they will include it in their article (if the results support the argument) or mention it briefly as a pretest or even discreetly discard it (if the results do not support the argument). For experiments using student labs or MTurk or similar online panels, I propose that we need norms linked to sample sizes. For example, perhaps any study with up to 15 respondents per cell should be a pilot study, such that researchers can conduct an unlimited number of them, but they do not appear as full-fledged experiments. Rather, experiments might require a sample size of at least 30 respondents per cell, and every experiment conducted should be reported, even if they do not provide the expected results. Such propositions would not apply to populations for which samples are costly or very hard to get, such as an f-MRI (functional magnetic resonance imagery) study, an experiment run with CEOs, or an investigation of consumers older than 90 years. Nor does it apply to cases in which the population of interest is very small, such that the sample covers all or most of it (e.g., all doctors with a specific specialization and qualification).

2.3. Avoiding “best of” and “verification bias” tactics

It is totally unacceptable for researchers to fabricate full data sets of any sort, whether by typing numbers into an Excel file and pretending they were obtained from experiments (Bhattacharjee, 2013) or by any other method, such as having simulation software create a multivariate data set with predefined relationships across variables and then pretending they are survey data. The Stapel (Levelt Committee et al., 2012) and Smesters (Simonsohn, 2013, pp.18 seq.) scandals at least (or at last?) should convince colleagues who could be tempted to engage in such forgeries that there is a high risk that they will be detected during the review stage or later, due to logical inconsistencies or implausible patterns in the fabricated data set.

But my concerns are not limited to such extreme cases. Respect for the data demands addressing and avoiding several forms of unethical behavior that hide or manipulate part of the collected data. In these cases, whether obvious or subtle, a discrepancy arises between the actual data and the analyzed data. The joint report of the three Committees investigating the Stapel case (Levelt Committee et al., 2012) provides an impressive series of examples:

1. Multiplying statistical tests to increase the chances that at least one is significant, then reporting this significant test selectively. This forgery requires collecting multiple measures of dependent or mediating variables and reporting only one, a subset, or a well-chosen combination. Another method collects data on multiple experimental conditions, manipulations, moderators, or product categories and reports only some of them.

2. Checking, after the collection of each observation, whether the result is significant (at 5%) and then stopping the data collection immediately, for fear that the result might no longer be significant after additional observations.

3. Comparing an experimental group from one experiment with the control group of another experiment, because the control group in the first experiment did not produce the desired contrast.

4. Removing experimental conditions or subgroups of respondents after the results are known.

5. Merging data from multiple experiments without mentioning it, to increase the number of respondents and arrive at significant results. This fraud should not be confused with a meta-analytic approach, which clearly acknowledges that it combines results from multiple studies.

6. Reporting reliabilities in a misleading manner (e.g., unreported values, erroneous values, values computed on subsamples, different ad hoc selections of items for the same scale in different studies, reference to a standard scale while using a nonstandard form).

7. Erroneous reports of p values.

8. Adding fictitious observations to those actually collected or selectively modifying the values of certain variables.

9. Replacing missing data with estimated data, without mentioning it.

A recent series of interesting articles has discussed such statistical manipulations: Simmons et al. (2011) offer the telling title “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant,” and Wagenmakers, Wetzels, Borsboom, and van der Maas (2011) discuss the dangers of an exploratory approach and suggest guidelines for confirmatory research. Finally, the Levelt Committee et al. (2012, p. 53) observe that journal reviewers encourage or even require some of these behaviors, such as removing certain experimental conditions. An International Journal of Research in Marketing (IJRM) editorial by Goldbergen and Muller (2012) provides an excellent example of the requirements that authors should follow to ensure integrity and enable better checks by the review team, such as making their data publicly available, reporting all measures and conditions, presenting results with and without covariates and with and without outliers, and explaining their choice of sample sizes. A JCR editorial (Luce, McGill, & Peracchio, 2012) warns against “best of” tactics and stresses the importance of full, present, and permanent disclosure, so that other researchers can understand fully how the research was conducted; their data and material also should be preserved for possible future re-investigation. Similarly, Marketing Science (Desai, 2013) now asks authors to submit data sets and estimation codes, to ensure that the research is replicable, while allowing for exceptions in well-specified cases (protected or compiled data sets, big data), as determined by the editor.

2.4. Better control by coauthors

Apparently, several of Stapel’s coauthors satisfied themselves with the story told by their colleague, namely, that he had collected, coded, and analyzed the data (with perfect results!). These coauthors never checked the original questionnaires and often did not ask for a copy of the data file or the computer output. The aftermath of the Stapel scandal has demonstrated that all coauthors of an article have an interest in and responsibility to check for possible ethical problems. If an article is

---

\(^6\) Among the 72 experimental articles that appeared in Vol. 39 of Journal of Consumer Research, 90% used student samples, 26% used online panels with minimal statistics on respondents’ age and gender, and 14% used online panels with no indication of demographics.

\(^7\) Amazon’s Mechanical Turk supports social science experiments run online. The choice of the name “Mechanical Turk” is rather strange; the original eighteenth-century “Mechanical Turk” was a fraud: a machine that apparently played chess well but actually was hiding a small human player.

\(^8\) According to Wikipedia (2013), “the cost of MTurk is considerably lower than other means of conducting surveys, with workers willing to complete tasks for less than half the US minimum wage.”
withdrawn or retracted, all coauthors suffer from the reduction in their publication record, as well as from suspicions, possibly unwarranted, due to their association with a tainted colleague, an “element of stigmatization that may persist long into their future career” (Levell Committee et al., 2012, p. 34).

Data collection and data analysis are too important to be left to a single data analyst. My recommendation here is simple: coauthors should organize systematic, reciprocal checking of their data collection, manipulation, and analysis. If a possible problem exists, it is much better to identify it and find a solution among the author team, rather than to wait for reviewers to expose it. Common practice by good market research companies also offers two simple, exemplary recommendations: (1) Always give precise details about the data collection, including respondents, dates, and locations, to enable subsequent controls in the lab or with the online panel and (2) maintain, in some dropbox, the data, programs, and output so that they remain available for random verifications by any coauthors.

Without denying the benefits of a division of labor among coauthors, according to each person’s specific competence, I believe that independent of possible ethical problems, there are benefits of involving coauthors in data decisions at all stages (i.e., developing the manipulations and instruments, running pilot tests, deciding on possible eliminations of observations, choosing possible transformations, even analyzing the data). The cost of having two coauthors work in parallel during each stage is minimal compared with the potential benefits of avoiding errors and exploring possible variants in the research process—not to mention improving the consistency across the conceptual framework, the data collection, the analysis, and the final write-up of the findings.

3. Discrepancies between experimental data sets and instances of the phenomena

In recent years, we have seen an increase in the frequency of experimental research in the consumer behavior field, with more articles using experimental approaches (i.e., 86% of the 84 articles published in Vol. 39 of JCR, more experiments per article (in JCR, 3.6 in 2010 vs. 1.7 in 1990), and more complex interactions per experiment (95% articles with two-way interactions and 33% with three-way in 2009 vs. 27% and 13%, respectively, in 1979 in JCR; Hamilton, 2012). In this section, I discuss four potential discrepancies between the data in an experiment and non-experimental instances of the phenomenon or underlying processes.

3.1. Constraints of experimental myopia

Whatever the positive benefits of experimental research, it remains a fundamentally myopic method, as used for consumer behavior studies.

3.1.1. Forward and backward myopia

Experimental research is almost always myopic in a longitudinal sense, because consumer behavior researchers study short-term reactions to short-term manipulations. That is, in most cases, respondents are exposed for a few minutes to a manipulation, then provide immediate reactions—sometimes after a “filler task” that lasts less than one hour. In better but rare cases, an effect might be measured a week or a month later. Mechanical Turk and similar online panels intensify this concern, in that one of the reasons for their attractiveness is their ability to deliver results in a few days, sometimes overnight. In contrast with medical or educational research, as well as unlike customer relationship management professionals, as consumer behavior researchers, we almost never apply manipulations and assess results over long periods, such as several weeks, months, or years (for an exception see Townsend & Liu, 2012). Thus we remain restricted to short-run causes and short-run consequences. Yet it is essential, when considering a phenomenon or theoretical process, to assess whether it is transitory or durable, and whether the first reaction to a stimulus triggers a subsequent, compensating, negative mediating mechanism that might counteract the initial reaction (e.g., a higher immediate clicking probability leads to lower clicking probabilities later; an icy road prompts more careful driving).

Consumers are long-lived. They receive repeated exposures to situations and stimuli. Over very long periods, often their entire lifetimes, they make repeated decisions, develop knowledge of and attitudes toward categories and brands, establish habits, and build heuristics. With the constraints of lab experiments, respondents do not have the time they normally would take. Instead, they have to develop new solutions to new problems immediately, in contrast with their actions when choosing, for the nth time, a brand of toothpaste, a perfume, a car, a radio station, or a movie. Decision processes in short-run experiments thus are not necessarily representative of consumers’ long-run decision processes.

Furthermore, short-term constraints force researchers to transpose real-life marketing stimuli and consumer responses into a miniaturized form that fits in the ephemeral time frame of an experiment. Studying long-run phenomena that cannot fit in such time frame should not be rejected by editors or abandoned by researchers. What if medical researchers studied the factors that cause cancer only in a short time frame?

3.1.2. Lateral myopia

Experimental myopia is also lateral: important consumer behavior problems simply are beyond the reach of experimental research using student subjects (which, as I noted previously, were used by 90% of the experimental articles published in Vol. 39 of JCR). Is it possible to extend results obtained with student samples to other samples? In a fundamental reference by Sears (1986) and a classical meta-analysis by Peterson (2001, p. 450), the recommendation is for caution when attempting such an extension, with an emphasis on the importance of “replicating research based on college student subjects with nonstudent subjects before attempting any generalizations” (see also Henrich, Heine, & Norenzayan, 2010; Henry, 2008; Hooge, Stolle, Maheo, & Vissers, 2010; and for a good example of non-student samples Volckner & Sattler, 2007). Three specific limitations seem critical.

First, there are limits to what can be manipulated. Consider some major changes in consumer behavior worldwide. In advanced countries, financial difficulties due to substantial unemployment rates force changes to consumption; in emerging countries, many consumers change their behavior when they benefit from markedly increased economic resources or move from a traditional village to a large city. Researchers might manipulate student subjects’ perceptions of wealth, by modifying the scale on which they report the balance of their bank account, and then assess the impact of this manipulation on some hypothetical (or even actual laboratory-based) consumption behavior. But is there any link between the processes the students undertake and those imposed by deep and enduring life changes? Similarly, changes in consumption entailed by deep shifts in family structures seem hard to study.

Second, we know that a continuous, regular decline occurs for processing-intensive tasks (e.g., speed of processing, working memory, and long-term memory) beginning when people are in their 20s and continuing through old age, whereas verbal knowledge increases across the life span (Park, Hedden, Davidson, Lautenschlager, & Smith, 2002). If an experiment with undergraduates reveals a phenomenon and underlying processes, can we generalize these results to consumers with lower abilities—cognitive or otherwise? To consumers outside of the Western world? To poorly educated consumers? Henrich et al. (2010, p. 61) caution about depending too much on samples drawn from “Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies,” because in many domains, the members of these societies are “among the least representative populations one could find for generalizing about humans.”
I do not mean to suggest a ban on experiments with students. Instead, I recommend that we take care not to restrict our view of the world to this small, carefully insulated cell. A research priority should be to show that results equally apply to them as they do to people in other geographical regions and cultures. Imagine the obverse: that sports researchers conducted experiments on the impact of different training schedules on performance only with participants aged 60 years and above? Would coaching practitioners conclude that the results probably apply to teenagers? The limitations entailed by student subjects should be especially bothersome for marketing researchers who, compared with social psychologists who study human behavior in general, are more interested in the behavior of specific consumer subpopulations (segments), defined according to age, education, income, cultural background, and so forth. From a marketing perspective, exploring the impact of such variables is much more important than imposing a simple statistical control of covariates.

Furthermore, experimental research on students is especially myopic in a domain with intense implications: the development of public policy recommendations for the rapidly growing elderly population, who suffer from decreases in multiple abilities. This population needs protection and assistance to take full advantage of consumption opportunities, especially those offered by new technology. To develop such policy recommendations, we need to study, using experiments and other approaches, older subjects who experience the complex syndromes that characterize this cohort.

Inspiration is available in the medical research field; we should use more alternative methodological approaches, such as real-life quasi-experiments, meta-analyses, epidemiological studies, econometric analyses of behavioral or survey data, and so on. If medicine had adopted the preponderance of experiments that appear in consumer behavior research, it would have been impossible to determine the impact of asbestos or cigarettes on cancer, because exposures to these carcinogens does not result from a random assignment of respondents (nor, returning to my first point in this section, do the effects arise after brief exposures).

Third, on a pedestrian level, when researchers use MTurk or a similar online panel, they should offer a precise description of the recruitment process, resulting demographics, and controls on basic qualifications, such as a good understanding of the questionnaire language.

3.2. Replications

As a Roman legal adage cautions, “one witness is no witness.” Respect for the data means we cannot accept a newly described phenomenon or process on the basis of a single, unique piece of empirical evidence in a specific setting with one operationalization—that is, a “singleton” (Ioannidis, 2005, 2012; Roediger, 2012). Yet several controversial examples come to mind, in which articles describe a new, spectacular effect that proves hard to replicate. In two famous cases, Bargh, Chen, and Burrows (1996) showed that priming undergraduates with the concept of “old age” led them to walk more slowly when exiting the lab, but Doyen, Klein, Pichon, and Cleeremans (2012) could not replicate the experiment identically. Bern (2011) asserted, among other “psi” processes, a “retroactive facilitation of recall” (pp. 419–20). Yong (2012) provides a clear description of these controversies.

In contrast with medical research, journals in our field have not been keen to publish replications, a preference of revived concern recently. As indicated by the title of a recent special section on replicability in Perspectives in Psychological Science (Pashler & Wagenmakers, 2012), it implies a crisis of confidence. Before turning to my recommendations, I note that this special section is insightful—especially Ioannidis’s (2012) contribution, as well as the useful in fine remarks by Albers (2012) on the difficulty of reproducing published results; I also commiserate with Evanschitzky, Baumgarth, Hubbard, and Armstrong (2007) and Makel, Plucker, and Hegarty (2012) about the extremely low rate of replication research in our domain and applaud the initiative of IJRM, which opened a new “Replication Corner” to encourage its publication (http://www.journals.elsevier.com/international-journal-of-research-in-marketing/news/ijrm-replication-corner-structure-and-process/).

We need two types of replications to answer two different questions. First, to confirm that outcomes are not due to random chance, it is necessary to obtain the same results in an identical replication, run in another lab by another researcher. (In turn, the original article must provide a full disclosure of the original experiment.) No less an authority than Kahneman (2012) suggests, for the specific case of priming research, organizing a chain of labs that can perform identical replications of their respective experiments. Second, we need “conceptual” replications in markedly different settings. As my linear algebra professor at MIT once said, “Anyone who describes a general theory and then provides a single application is cheating you” (Abelson, 1974, private communication). Articles that test complex interactions between several conceptual variables using a single operationalization of each concept and samples from the same subject pool are suspect. Replications should differ from the initial study by sampling another population (as discussed previously), relying on different research methods, or using different operationalizations of the conceptual variables. It is especially important to provide a replication for a new phenomenon that has first been discovered among a sample of students.

Therefore, I propose requirements that mandate that authors who first describe a new phenomenon with a student or MTurk sample provide a conceptual replication in a markedly different setting, or a “self-replication.” (This proposition does not apply to cases in which samples are very costly or hard to get, as outlined in Section 2.2.) As an additional benefit, this requirement would force authors to delineate more precisely the concepts they study by developing two different operationalizations; the risk of confounds decreases with more distinct operationalizations. This proposition concurs with Winer’s (1999) plea to combine lab-based experimental research with modeling approaches based on scanner or other types of data. It also reflects Ehrenberg’s lifelong insistence on “empirical generalizations,” that is, regularities in results obtained in various settings (e.g., Uncles, Ehrenberg, & Hammond, 1995).

If an article follows up on previous research, respect for the data means that it should start with a replication of the previous evidence, as “Study Zero”. The replication could be identical if the original article contained a replication in a different setting or conceptual if the original article used a single setting, procedure, population, or operational definition—and especially if that prior article offers the sole previous evidence of the phenomenon. The description of the replication in the new article could be very brief or even relegated to a web appendix, but I consider it important that the replication takes place. This recommendation is in the interest of the new investigator: there is no point in embarking on a project that builds on previous work if the phenomenon investigated cannot be replicated.

Exceptions to these requests for replications could be granted at the discretion of editors, such as when the problem being tackled is urgent or has important policy implications (e.g., protecting children, older consumers, addicted persons, or drivers); the contribution features a new methodology (e.g., the data set serves only as an example); the data collection is exceptionally time consuming or costly (e.g., a business-to-business study with multiple informants in many companies that merges several databases); extensive databases are involved (e.g., scanner panels, store censuses); or because confidentiality or privacy demands it.

---

9 Testis unus testis nullus.
3.3. Considering the phenomenon and its underlying processes in their entirety

When researching a phenomenon, it needs to be considered in its entirety, rather than in a narrowly limited setting or for only a specific data set. In particular, it is far more convincing to demonstrate a phenomenon or process with multiple operational manipulations of a conceptual independent variable (IV) or moderator (e.g., Wan & Rucker, 2013, who use three different manipulations of high and low confidence levels) than with the same operational manipulation across all studies in an article. A single operational definition runs a much higher risk of confusing the conceptual variable of interest with other conceptual variables, and the results easily could be due to some other aspect of the manipulation. If an operational definition makes its début in the article or has been seldom used previously, the author should discuss extensively why it is valid and reliable. Such efforts are less necessary when the operational definition has been extensively validated by previous research (e.g., a moderator assessed by a test included in Wechsler’s Adult Intelligence test).

Multiple measures of the dependent variable are helpful, but the ultimate goal should be to verify that the measures are reasonably correlated, not just pick the measure that best supports the authors’ argument after the results emerge. Implemented measures should be reported, along with their links. Again, unless the measures have been extensively used previously, the authors should justify why each one is valid. Although discussion of the choice of the mediators and moderators contained in a study is common, it also would be interesting to add a discussion of mediators and moderators that might have been but were not included.

Many experimental articles also report the results related only to the variables they manipulate. To avoid what econometricians call a “specification error” (i.e., omitting an explanatory variable, which can lead to biased and inconsistent estimators), researchers should not restrict themselves only to the variables in which they are interested and manipulate; instead, we should analyze and report, as much as possible, the effect of all the variables that might affect the dependent variable. Specifically, authors should report the impact of natural (measured) covariates or moderators such as gender, age, and level of education, including possible interactions with manipulated variables. It is more efficient statistically to control the impact of such characteristics by including relevant variables in the equations than simply to assume that their impact is part of the random term.

3.4. Size and strength of phenomena

In published articles, the existence of an effect is too often evidenced by just a significance test (t or F). This is not enough. Researchers should also report the strength and size of the effect in absolute terms: How large is the adjusted $R^2$ or $\omega^2$? How strong is the elasticity? In addition, the reports should include comparisons with the effect of covariates. Is priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of local advertising? If priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of local advertising? If priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of local advertising? If priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of local advertising? If priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of local advertising?

The effect of all the variables that might affect the dependent variable. Specifically, authors should report the impact of natural (measured) covariates or moderators such as gender, age, and level of education, including possible interactions with manipulated variables. It is more efficient statistically to control the impact of such characteristics by including relevant variables in the equations than simply to assume that their impact is part of the random term.

3.4. Size and strength of phenomena

In published articles, the existence of an effect is too often evidenced by just a significance test (t or F). This is not enough. Researchers should also report the strength and size of the effect in absolute terms: How large is the adjusted $R^2$ or $\omega^2$? How strong is the elasticity? In addition, the reports should include comparisons with the effect of covariates. Is the elasticity of increased shelf space greater or smaller than the elasticity of local advertising? If priming subjects with old age leads them to walk more slowly, is the impact stronger or weaker than the impact of actual physiological age? Reports of effect strengths and sizes also should be standardized, allowing for comparisons across studies and meta-analyses, such as in the form of elasticity coefficients (Albers, 2012).

At the same time, editors need to consider the strength of the demonstrated effects as an important criterion to evaluate a research contribution. Respect for data implies that we learn more by revealing a strong effect than a weak effect, or a high elasticity rather than a small elasticity. We cannot limit ourselves to qualitative evidence. Even if an effect has strong qualitative support, it continues to be important to learn when it is strong and when it is weak (e.g., Lodish et al., 1995, who study advertising effects, and their follow-ups). After using different operationalizations of a conceptual variable, it is important to indicate whether they have the same impact or which is stronger. Practical implications for government authorities and companies are more likely when the effect is important. Asking for strong effects also reduces the risk of accepting an article with false positives.

Respect for the data requires caution about another tendency too, namely, the habit of review teams to ask for “non-obvious” results. These hindsight effects (Bernstein, Erdfelder, Meltzoff, Peria, & Loftus, 2011; Slovic & Fischhoff, 1977) could lead to a rejection, because the review team thinks, once the results are known, that the hypothesized relationships and results are “obvious,” in that they correspond to intuition. But without having read the manuscript, their intuition might have been different. A feeling of “obviousness” is not data. It certainly is reasonable to reject an article if its hypotheses and results are very similar to what has been demonstrated by previous research and offer few or marginal insights. But this situation differs from cases in which, absent previous research, the main argument for rejection is that the results appear, after the fact, intuitive or obvious to the review team.

These criticisms of experimental research also should not blind us to the limits of other data collection methods. Surveys suffer from problems (Rindfleisch, Malter, Ganesan, & Moorman, 2008), such as low response rates; disagreements among multiple respondents from the same organization, which cast doubt on the reliability of surveys relying on single informants; or response styles and halo effects. Marketing research relies on multiple forms of data, including scanner data from store panels or censuses, online or mobile phone records, social media data, consumer panels, archival data from companies, field and quasi-experiments, and time series. As these examples show, we must respect data, whatever form they take. My focus on experimental research here is due to space constraints.

4. Let the data speak, beyond predefined scripts

Published consumer behavior research reveals the predominance of hypothetico-deductive approaches, presenting experimental results based on preexisting theory. Respect for the data should make us ready to accept unexpected results too. We cannot risk the circular trap of refusing to accept results if they are not based on a preliminary theory. This important question already has been addressed by several senior colleagues. Alba (2012, p.985), “in defense of bumbling,” states that it is “not illegitimate to engage in abduction ... or to acknowledge that an if-then statement can be valuable even if the intervening causal link has not or cannot be identified.” Lynch (2011, pp.1 and 4) also encourages researchers to “more often look to the substantive domain as inspiration for our research” and “start with the consumer phenomenon and then try to explain it rather than always starting with concepts in the literature and then thinking of where they might apply.” According to Lynch, Alba, Krishna, Morwitz, and Gürhan-Canli (2012), we should be open to both non-deductive and deductive routes. Park (2012) also argues for “incomplete” or “cute” research, which features novel, interesting empirical findings, even if the authors have not identified the underlying processes. As Albers (2012, p. 121) writes, “reviewers should accept more studies that are descriptive and not necessarily test a theory.” In this section, I offer three brief, personal examples of a non-deductive approach.

4.1. Revealing multidimensionality

To measure a conceptual variable, the empirical need for a multidimensional measure may reveal a multifaceted conceptual variable. When Jean-Noël Kapferer and I embarked on developing an empirical measure of involvement (Laurent & Kapferer, 1985), we thought it would be a unidimensional scale. Through an iterative process, tacking across qualitative interviews, item wording, data collections, and analyses, we concluded instead that it was necessary to measure consumer involvement profiles (pleasure value, symbolic value, risk importance, probability of error) rather than a single unidimensional involvement.
score. This realization led us to identify multiple conceptual facets of involvement.

An overreliance on confirmatory statistical analyses may limit the rich potential benefits of exploration. It is dangerous to try to confirm possible theoretical relationships involving a construct designated by a single word without being sure that that very word denotes a single, unidimensional construct. An exploratory factor analysis approach with multiple items that reportedly measure a conceptual variable may reveal whether there is a unique concept aligned with that name. In a confirmatory approach, imposing a priori unidimensionality on a concept instead may lead to the elimination of items that do not fit the predetermined schema, even though they actually represent an important facet of the concept. Similarly, it is useful to test alternative manipulations of a conceptual variable with the same checks (though we should not apply confirmatory methods designed for reflective constructs to formative

4.3. Identifying complex functional forms

When an empirical scatter plot indicates that the relationship between two numerical variables, x and y, is not linear, respect for the data implies the need to identify the appropriate functional form. I refer here to cases in which the full set of observations diverges from a linear relationship, not when one or a few observations truly stand out from an otherwise linear relationship (e.g., observing whether outliers with similar values of the explanatory variable tend to have similar values of residuals or “deleted residuals”¹¹). To identify the functional form, two approaches are viable. If the research project focuses on a specific data set, the researcher can use the approach described by Albers (2012, pp.112–115) to identify a nonlinear function of x that fits well with the values of y. First, consider the functional forms that respect the logical constraints (e.g., include diminishing returns at least above certain levels). Second, use an exploratory approach that computes the moving averages of y and observes how they change as a function of x (visual inspection). If the project instead focuses on many parallel data sets (e.g., different categories, different countries), the goal is to find a single functional form that fits all data sets (allowing at most for a change in parameters) while respecting the logical constraints. Tukey (1977, chapters 5 and 6), in the same influential book I cited previously, not only introduced the box plot but also suggested ways to try to identify “re-expressions” of either x or y, or both, that “straighten out the plot” (the title of his Chapter 6) and identify a transformation of the variables that allows the relationship between the transformed variables to become linear. Identifying the re-expression that straightens the plot should reveal the underlying structure of the relationship. Obvious examples are exponential growth and decline, for which the plot of y against time (x) is nonlinear, but the plot of the logarithm of y against time is linear, which indicates the multiplicative impact of time. Tukey again suggests a variety of possible re-expressions: logarithms, reciprocals, and powers. In my own research (Laurent, Kapferer, & Roussel, 1995), after observing that in each of 39 different product categories, there was an almost perfect linearizing a relationship still demands caution though, especially with regard to the distribution of residuals (and therefore random error terms), around both the original nonlinear and the linearized relationships: Is it Gaussian? Is it homoskedastic? What are the consequences for estimation?

¹¹ In addition to traditional regression residuals, standard software packages (e.g., SPSS) offer more complete “influence diagnostics” (Belsley, Kuh, &Welsh, 1980) for each observation in the data set, such as its deleted residual (prediction error for one observation if the regression is estimated without that observation), dft (how much the estimated beta coefficient would be changed if this observation were eliminated), and so on. As the name indicates, these tools can detect which observations have strong influences on the regression results. Such observations should not just be dismissed, because they may result from nonlinear relationships.
Overall, respect for the data recommends letting the data speak, even in the absence of a predefined script, as in these three examples.

5. Conclusion

As I wrote in the introduction, the lessons to be derived from recent scandals should not be limited to cautiousness about utter forgery. We need to be more broadly respectful of our data. Therefore, in Table 1 I summarize this discussion, in terms of specific actions that I believe researchers, editors, and journals should (or should not) undertake. Beyond these specific recommendations, I conclude with three general pleas.

First, respect our data sets, collected following our design and under our control. Data cannot be subordinate to researchers. We are not free to mutilate our data at will and without reporting our actions. Observations may be set aside only for good, specific reasons—not just because they take extreme values. Such actions also should be rare, assuming that our experiments and measures have been well designed. Nor may we hide experiments in file drawers when they do not produce the expected results. Transparency is essential. At the same time, editors and reviewers must stop requiring fairy-tale reports in which each experiment works as expected.

Second, respect vast data, beyond our narrow experimental data sets, and address possible discrepancies among them. Is it reasonable to exclude alternative, non-experimental methods that have proved so useful in medical fields, such as quasi-experiments, meta-analyses, epidemiological studies, and econometric analyses of behavioral or survey data? Is it reasonable to rely almost exclusively on student samples or samples from Mechanical Turk or similar online panels? Is it really reasonable to assume that the vast data available outside the lab are inferior to data collected in that narrow, insulated cell? Never in the history of consumption research have we seen so many diverse groups of consumers undergoing so many structural changes worldwide, due to massive income shifts in both developing and developed countries, substantial increases in the number of older consumers, and worldwide innovations in distribution and information networks. How can we remain blind to these developments and keep looking only at our WEIRD labs?

Third, respect data that tell us something unexpected. In contrast with Stapel’s claim, cited in the introduction, when an experiment does not give us what we are looking for, we need to question the theory as much as the data. Without completely abandoning experiments and the hypothetico-deductive approach, we need to be more open to discovering new, unexpected relationships from data collected with alternative methods.

Ultimately, we should respect the data because we should respect the phenomena and underlying processes we study. We need a balance among the researcher, the theory, and that which is being researched. The latter should never be subordinate, and nor should data ever appear in a subordinate role.

Acknowledgments

I thank Soenke Albers, Gary Lilien, Berend Wierenga, and Eitan Muller for their extremely useful comments and suggestions on previous versions of this article. I also thank Jean-Louis Chandon, Joe Lajos, Raphaëlle Lambert-Pandraud, Don Lehmann, and Mark Uncles for their suggestions. This article benefited from discussions at two panels, held at HEC Paris in September 2012 and ACR in October 2012. Ultimately though, the opinions and recommendations suggested herein are solely my own responsibility.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijresmar.2013.07.003.

References

Do not mutilate data sets

1. Do not hide experiments that did not “work” (did not produce the expected results). Specify and, if possible, develop and test hypotheses for why the experiments did not produce the expected results.

2. Clearly decide, before collecting data, whether they will serve a pilot test or an experiment.

3. When discard alleged “outliers,” mention them briefly in the text and justify the choice, perhaps in a web appendix: Give precise reasons and the raw values of each discarded observation, along with a description of the distributions before and after removing these outliers (including estimated within-group variance).

4. Do not remove alleged “outliers” solely because they locate beyond the whiskers in a boxplot (i.e., more than 1.5 IQR beyond the quartiles).

5. Use different procedures to handle alleged outliers in the same article if the distributions of the dependent variables differ across studies.

6. Check the distribution of the dependent variables before performing an ANOVA. If far from Gaussian, consider a non-parametric analysis or re-express the variable to bring it closer to a Gaussian distribution before performing an ANOVA.

Avoid discrepancies between experimental data sets and other instances of the phenomenon or process

7. When revealing a new phenomenon or new process, provide a conceptual replication with a different operationalization in a different setting.

8. When following up on previous research, begin with a replication of the previous results.

9. Fully disclose research procedures to enable replications by other researchers.

10. Avoid “best of” or “verification bias” tactics (Levitt Committee et al., 2012), as stated in Section 2.3 of this article.

11. Use multiple manipulations and measures unless the manipulation or measure has been extensively validated by previous research. Otherwise, justify why they are valid and reliable operationalizations of the conceptual variables. Report all the measures of a variable and check for convergence.

12. Do not restrict subject pools to undergraduates students or Mechanical Turk samples. In any given article, draw samples from a reasonable variety of populations, rather than from a single student population or MTurk service; if possible, draw them from different geographical regions and cultures.

13. If using MTurk, control for and report demographic descriptors and participants understanding of the language.

Let the data speak outside of their predefined script

14. Be open to do research that does not follow the hypothetico-deductive approach.

15. Be open to accept results even in the absence of anticipated theory, unexpected discoveries, or unpredicted relationships.

16. Do not discard results because they do not fit predefined schemas.

17. Beyond lab experiments, be open to alternative data collection methods, such as real-life quasi-experiments, meta-analyses, epidemiological studies, and econometric analyses of behavioral or survey data.

18. Research important topics in consumer behavior (from public policy, theoretical, or managerial points of view), even if they cannot be studied in laboratory experiments using students or online panels.

19. Be cautious about hindsight bias when tempted to reject “obvious” results.

20. Avoid specification errors by including all available variables (and interactions when appropriate) in analyses.

Miscellaneous

21. Report effects strength (e.g., adjusted R²) and size (e.g., elasticity coefficient).

Consider the strength of the effects as an important criterion when evaluating an article.

22. Arrange for coauthors to check one another’s work at every stage.


Consumer responses to variety in product bundles: The moderating role of evaluation mode

Xia Wang a,1, Luping Sun a, Hean Tat Keh b,*

a Department of Marketing, Guanghua School of Management, Peking University, Beijing 100871, PR China
b Department of Marketing, Faculty of Business and Economics, Monash University, PO Box 197, Caulfield East, VIC 3145, Australia

ARTICLE INFO

Article history:
First received in 13, October 2011 and was under review for 8 months
Available online 20 June 2013

Guest Editor: Marnik G. Dekimpe
Area Editor: Tom Meyvis

Keywords:
Variety
Product bundles
Evaluation mode
Satiation

ABSTRACT

This article examines the moderating effect of evaluation mode on consumer responses to variety in product bundles. Study 1 finds that consumer preference for the variety bundle (relative to the non-variety bundle) is higher in the joint evaluation mode than in the separate evaluation mode. Study 2 provides evidence that the increased preference for the variety bundle in joint evaluation is driven by the activation of concerns about satiation. Specifically, we find that both the quantity of items in and the category of the non-variety bundle influence consumer concern for satiation and the evaluation of the variety bundle. Study 3 further examines the proposed mechanism by manipulating the information associated with repetition and finds that associating repetition with loyalty (vs. satiation) eliminates the moderating effect of evaluation mode on the preference for variety. We discuss the findings and their implications for marketers.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

During planned shopping trips (as opposed to impulse purchases), consumers often buy more than one unit of a given product category, particularly for convenience goods that have to be replenished frequently (e.g., snacks and yogurt). Correspondingly, manufacturers and retailers see the benefit of bundling multiple units of the product into a single package to facilitate consumer purchases (Dempsey, 1959; Simonson, 1999). In so doing, a key decision pertains to whether manufacturers and retailers should offer non-variety bundles (e.g., six cups of yogurt that all have the same flavor) or variety bundles (e.g., six cups of yogurt where each comes in a different flavor).

The variety-seeking literature informs us that consumers are generally motivated to seek variety to increase stimulation, reduce satiety, and withstand future uncertainty (Kahn, 1995; McAlister & Pessemier, 1982; Simonson, 1990). In particular, research on diversification bias suggests that when people have to make simultaneous choices for future consumption (e.g., choose now which of six snacks to consume in the next four weeks), they tend to seek more variety than when they make sequential choices (e.g., choose once a week which of six snacks to consume that week for four weeks) or separate choices immediately preceding consumption (Read & Loewenstein, 1995; Simonson, 1990). Thus, in practice, it is not surprising to find that manufacturers and retailers often endow product bundles with some variety (Anonymous, 2006; Howell, 2000).

However, from the consumer’s perspective, there is also a potential downside of choosing the product bundle with variety; by definition, the bundle with variety implies that consumers would have fewer quantities of their most-preferred items and need to compromise on the less favorable ones. Given both the potential benefit and downside, we are interested in understanding how consumers evaluate and respond to variety in product bundles. As consumers’ variety-seeking behavior is usually influenced by contextual factors (e.g., Ratner & Kahn, 2002), we adopt the context-dependent approach to understand their evaluation of variety in product bundles. In particular, we examine one of the most common contexts for bundle evaluation—the evaluation mode. When the bundle with variety is evaluated by itself (i.e., separate evaluation), the presence of variety merely implies the aggregation of different items, and consumers’ evaluation of the bundle would depend on their average preference for the various items in the bundle. However, when the bundles with and without variety are evaluated simultaneously (i.e., joint evaluation), we posit that the direct comparison will activate consumers’ concern for satiation in the non-variety bundle, making the variety bundle more desirable. That is, the absence (vs. the presence) of the non-variety bundle may contextually influence the value derived from the variety bundle. To examine this proposition, we conduct three studies to test the moderating effect of evaluation mode on consumers’ preference for variety and the underlying mechanism.
of their concern for satiation. Exploring this issue will not only enrich our understanding of consumer responses to variety in product bundles, but also provide relevant insights for manufacturers and retailers to improve the marketing of these bundles.

The rest of the paper is organized as follows. We first review the relevant literature and develop the hypotheses. Then, we conduct three studies to test our hypotheses. Finally, we discuss the theoretical and managerial implications of our findings.

2. Theoretical background

2.1. Satiation and variety seeking

Past research shows that variety seeking, either as an inherent or context-motivated tendency, can influence consumer choices (Kahn, 1995; McAlister & Pessemier, 1982; Simonson, 1990). A fundamental motivation for consumers to seek variety is to avoid the satiation that comes from repeated consumption of an item (McAlister, 1982). Because consumers quickly become satiated with the attributes of a product, they desire products or experiences that offer a range of other attributes. Thus, switching among the choice alternatives allows consumers to reduce satiation. Other benefits that derive from variety seeking include enhancing stimulation (Galak, Redden, & Kruger, 2009; Inman, 2001), augmenting freedom and flexibility (Mogilner, Rudnick, & Iyengar, 2008), magnifying anticipated enjoyment (Kahn & Wansink, 2004), leaving a favorable and interesting memory (Ratner, Kahn, & Kahneman, 1999), and delivering a unique or interesting impression (Ratner & Kahn, 2002). Consequently, consumers are willing to tolerate less-preferred items for the sake of variety (Ratner et al., 1999).

In particular, variety seeking is most pronounced for products with hedonic attributes on which individuals quickly satiate (e.g., flavors), rather than non-hedonic attributes (e.g., brands; Inman, 2001). In addition, as consumers have the lay belief that they will satiate on favored items quickly, they tend to incorporate more variety for future consumption in simultaneous choices than in sequential choices (Simonson, 1990). Thus, when consumers think about the satiation that will result from repetition, they show diversification bias by incorporating more variety to counteract the anticipated physiological satiation (Read & Loewenstein, 1995). Notwithstanding these findings, other research indicates that consumers also seek variety in sequential choices (Ratner et al., 1999), given that repetition is associated with boredom and signals closed-mindedness, whereas variety seeking prevents satiation and signals open-mindedness (Ratner & Kahn, 2002).

Rather than being an inherent and stable physiological process, satiation can be constructed in the moment. For example, the memory of past consumption plays an important role in determining satiation (Redden & Galak, 2013). Specifically, recalling the variety in past consumption is helpful in accelerating the recovery from satiation (Galak et al., 2009). Furthermore, variety-seeking behaviors can even be triggered unconsciously by situational cues that activate the satiation-based (vs. preference-based) construal of the choice (Fishbach, Ratner, & Zhang, 2011).

Although variety is an important consideration when consumers make product choices (Kahn, 1995), the literature has little to say about how consumers evaluate variety in product bundles. Exploring this issue from the consumer’s perspective will extend our understanding about the value of variety in product bundles, as influenced by the anticipated satiation for bundle consumption. In addition, prior research finds that variety-seeking behavior is generally context dependent. For example, positive consumer mood (Kahn & Isen, 1993), public (vs. private) consumption occasion (Ratner & Kahn, 2002), and even irrelevant varied arrays (Maimaran & Wheeler, 2008) can influence individuals’ desire for variety. Accordingly, the current research seeks to examine whether consumers’ responses to variety in product bundles vary across different evaluation modes that have the potential to influence the construction of satiation.

2.2. Two modes of evaluation

According to Hsee (1996, 2000), all judgments and decisions are made in one of two basic modes—joint evaluation (JE) mode and separate evaluation (SE) mode. In the former condition, consumers are exposed to multiple objects simultaneously and evaluate these objects comparatively, while in the latter condition, they are exposed to only one object at a time and evaluate it in isolation. Thus, direct comparison is available only in the JE mode, which facilitates comparison of attributes that are intrinsically difficult to evaluate. Prior research indicates that the salience of product attributes would be quite different in the two evaluation modes; the easy-to-evaluate attributes are more salient in the SE mode, while the hard-to-evaluate attributes are more salient in the JE mode (Hsee, 2000; Kahneman & Thaler, 2006). Consequently, consumer preferences could be different or even reversed under different evaluation modes (for a review, see Hsee, Loewenstein, Blount, & Bazerman, 1999).

In this research, we extend the application of the evaluation mode to study consumer responses to variety in product bundles. We propose that the contextual factor of evaluation mode has the potential to influence consumers’ in-the-moment construction of satiation and, thus, their subsequent evaluation of variety in product bundles. Specifically, compared to separate evaluation, simultaneous exposure to the variety and non-variety bundles will trigger anticipated satiation for the non-variety bundle, leading consumers to favor the variety bundle. Exploring differential consumer responses to product bundles with and without variety as moderated by the evaluation mode can not only contribute new findings to the literature, but also provide meaningful implications toward marketing practice.

2.3. Present research

We propose that it is difficult for consumers to evaluate the value of variety in product bundles when a reference is unavailable (i.e., in the SE mode). However, when the bundle without variety is displayed alongside that with variety (i.e., in the JE mode), the anticipation of consuming the same items in the non-variety bundle is likely to trigger consumers’ concern for satiation and induce their preference for the variety bundle to counteract the potential satiation. That is, consumers’ aversion toward satiation from repeated consumption of the same items enhances their motivation to seek variety, which makes the variety bundle more desirable in the JE mode than in the SE mode. We test this central hypothesis in three studies.

Study 1 examines the baseline in influence of evaluation mode; that is, whether the presence of variety works better when the variety bundle is displayed by itself or when it is juxtaposed alongside a non-variety bundle. Study 2 explores the mechanism underlying the JE mode by varying the comparison non-variety bundle in terms of the quantity of items it has and its category. Through manipulating the information (satiation vs. loyalty) associated with repetition, study 3 further examines how the concern for satiation acts as a means of influencing the moderating effect of evaluation mode.

3. Study 1: The influence of evaluation mode on consumer responses to variety in product bundles

We propose that, in the SE mode, consumers perceive the product bundle as a combination of the items in it and that the value derived from variety in that bundle is limited. However, in the JE mode, the direct comparison between the variety and non-variety product bundles may activate consumers’ concern for satiation in the non-variety bundle, leading consumers to prefer the variety bundle.
3.1. Design and sample

3.1.1. Experimental design

In study 1, we used bundled potato chips as the stimuli, and conducted a pretest with 30 participants similar in profile to those in the main study to identify the five most popular flavors of potato chips. Based on the pretest, we developed a pair of stimuli, each containing five small bags of equal weight (i.e., 50 g/bag) that have the same price. The only difference between the two stimuli was the presence versus the absence of variety on the attribute of flavor. Specifically, stimulus A (i.e., the non-variety bundle) had five small bags in the most popular flavor (i.e., BBQ flavor), while stimulus B (i.e., the variety bundle) had five small bags that each came in a different flavor (i.e., BBQ, tomato, original, spicy, and chicken flavors). Then, we designed three conditions to test the effect of evaluation mode—SE for stimulus A, SE for stimulus B, and JE for both stimuli. In the first two conditions, the participants were exposed to only one stimulus (either A or B) and were required to evaluate it. In the third condition, the participants were exposed to both stimuli simultaneously and required to evaluate each of them. The order of the two stimuli was counterbalanced in the third condition.

3.1.2. Sample

Participants were recruited using an online advertisement on the bulletin board system of a large public university and were compensated for their involvement. The final sample had 104 participants (45.2% male, mean age = 22.1 years).

3.2. Procedure and measures

3.2.1. Procedure

The participants were randomly assigned to one of the three conditions. In each condition, they were first exposed to verbal and pictorial descriptions about the stimulus/stimuli, and then asked to indicate their bundle evaluations. Afterwards, they were asked to list any thoughts they had when evaluating the bundle(s). The participants were also asked to indicate their favorability for the flavor(s) they were exposed to and some demographic information. At the end, they were debriefed, compensated, and dismissed.

3.2.2. Measures

In this study, we measured consumer evaluations of the product bundles based on bundle favorability and purchase intention. Bundle favorability was measured on a 7-point scale (1 = not at all favorable; 7 = extremely favorable). Purchase intention was measured by the likelihood that the participants would purchase the product bundle, ranging from 0% to 100%. In addition, three items were used to measure perceived variety: “This product gives a lot of variety for me to enjoy,” “This product offers more ways to enjoy it” (for both items: 1 = strongly disagree; 7 = strongly agree), and “How much variety do you think there is in this product?” (1 = very little variety; 7 = very much variety) (Cronbach’s α = .91; Kahn & Wansink, 2004). Consumer favorability for the specific flavor(s) was measured on a 7-point scale (1 = not at all favorable; 7 = extremely favorable).

3.3. Results and discussion

We found no significant difference among the three conditions in terms of preference for the BBQ flavor (F(2, 101) = 1.15, p > .10). Among the five flavors, BBQ was consistently rated the highest across the three conditions, implying that the non-variety bundle (i.e., BBQ flavor only) should be the optimal choice for consumers. Interestingly, when the variety bundle was evaluated in the SE mode, the participants’ bundle favorability (M = 3.77) was lower than their average favorability for the five individual flavors (M = 4.58; F(1, 34) = 20.44, p < .001). In contrast, when the variety bundle was evaluated in the JE mode, the participants’ bundle favorability (M = 5.23) was significantly higher than their average favorability for the five flavors contained in it (M = 4.70; F(1, 34) = 6.02, p < .05). This finding implied that consumers derived additional value from variety, beyond the simple combination of different flavors, but only in the JE mode.

As shown in Fig. 1, in the SE mode, the favorability for the variety bundle was not significantly different from that for the non-variety bundle (Mvariety = 3.77 vs. Mnon-variety = 3.26, t(67) = 1.23, p > .10). However, in the JE mode, consumers’ favorability for the variety bundle was significantly higher than that for the non-variety bundle (Mvariety = 5.23 vs. Mnon-variety = 3.27, t(34) = 7.65, p < .001). The difference between the variety bundle and the non-variety bundle on favorability was significantly larger in the JE mode than in the SE mode (t(101) = 3.40, p < .001). The results for purchase intention revealed a similar pattern. Specifically, there was no significant difference between the variety bundle and the non-variety bundle in the SE mode (Mvariety = 41.29 vs. Mnon-variety = 35.91, t(67) = .83, p > .10), while the variety bundle was more likely to be purchased than the non-variety bundle in the JE mode (Mvariety = 67.29 vs. Mnon-variety = 22.94, t(34) = 7.83, p < .001). Notably, the difference in purchase intention between the variety bundle and the non-variety bundle was significantly larger in the JE mode than in the SE mode (t(101) = 4.54, p < .001). Taken together, these results indicated that consumer evaluation of the variety bundle (vs. the non-variety bundle) was moderated by the evaluation mode.

We further explored what led to differential consumer responses toward the two bundles under the two evaluation modes. We found that the participants’ favorability for the non-variety bundle did not differ across the two modes (MSE = 3.26 vs. MSE = 3.27, F(1, 67) < .001, p > .50); however, their favorability for the variety bundle was significantly higher in the JE mode (MJE = 3.77 vs. MSE = 5.23, F(1, 68) = 17.35, p < .001). More interestingly, compared with the SE mode, the participants’ purchase intention for the non-variety bundle in the JE mode was reduced (MJE = 35.91 vs. MSE = 22.94, F(1, 67) = 5.39, p < .05), while that for the variety bundle was increased (MJE = 41.29 vs. MSE = 67.29, F(1, 68) = 17.16, p < .001). Taken together, these results indicated that the primary driver of the moderating effect of evaluation mode was the favorable response toward the variety bundle in the JE mode, which was not apparent in the SE mode.

We also conducted a content analysis on the participants’ listed thoughts when evaluating the bundle(s). Two marketing doctoral students who were unaware of the research purpose classified the thoughts listed by the participants based on keywords. Specifically, the thoughts on “satiation” or “boring” were indicative of the participants’ concern for satiation. We found that 58% of the participants expressed concern for satiation when evaluating the bundles in the JE mode, which was significantly higher than when the participants separately evaluated either the variety bundle (14%; z = 3.26, p < .001) or the non-variety bundle (20%; z = 3.12, p < .01). This finding indicated that comparing the variety and non-variety bundles in the JE mode activated consumers’ concern for satiation, highlighting the role of variety in counteracting potential satiation from the non-variety bundle. Nonetheless, study 1 did not directly examine

\(^2\) When the variety bundle was evaluated in the SE mode, the participants’ favorability for each flavor was as follows: Mchicken = 4.91, Mtomato = 4.69, Mbottle = 4.63, Mtomato = 4.20, and Mchicken = 4.46. The average favorability for the five flavors was M = 4.58, which was higher than the bundle favorability (M = 3.77).

\(^3\) When the variety bundle was evaluated in the JE mode, the participants’ favorability for each flavor was as follows: Mchicken = 4.98, Mtomato = 4.92, Mbottle = 4.87, Mtomato = 4.46, and Mchicken = 4.27. The average favorability for the five flavors was M = 4.70, which was lower than the favorability for the variety bundle (M = 5.23).

\(^4\) Our analysis followed the procedure in Hsee (1996), whereby the evaluations of the two bundles in the JE condition came from the same participants.
the underlying mechanism of the concern for satiation, and we conducted study 2 to address this limitation.

4. Study 2: How the JE mode helps manifest the value of variety in product bundles

We propose that the concern for satiation triggered by the non-variety bundle leads consumers to appreciate the value of variety in the JE mode. Past research indicates that consumer satiation is mainly induced by repeated consumption of the same items (McAlister, 1982). Consumers anticipate and even overestimate their satiation when considering the same items for future consumption (Read & Loewenstein, 1995). Thus, in study 2, we manipulate consumer concern for satiation in future consumption by varying the quantity of items in the non-variety bundle. We posit that the non-variety bundle with a small quantity will be less effective in activating satiation and, therefore, be less likely to lead to increased consumer evaluation of the variety bundle in the JE mode. In contrast, when the non-variety bundle has a large quantity of items, the activated satiation will be relatively strong, thus leading to a higher evaluation of the variety bundle. However, we also expect that when satiation is activated by a non-variety bundle from a different (vs. same) category, the value of the variety bundle in counteracting the anticipated satiation will be reduced. Thus, the impact of a large (vs. small) quantity of items in the non-variety bundle on the evaluation of the variety bundle will be weaker when the two bundles are from different categories.

4.1. Design and sample

We used instant coffee as the target stimulus in study 2; the variety bundle contained 12 packs (15g/pack) of instant coffee, which came in four different flavors (i.e., Latte, Cappuccino, Mocha, and Black coffee). To manipulate the satiation triggered by the non-variety bundles in the JE mode, this study used a 2 (quantity of items in the comparison non-variety bundle: small vs. large) × 2 (product category of the non-variety bundle: coffee vs. cookie) between-subjects design, and we also added a SE condition for the variety bundle as a control. The category of the comparison non-variety bundle was either the same (i.e., coffee) or different (i.e., cookie) from the variety bundle. In terms of quantity, the small-quantity non-variety bundle had four packs, while the large-quantity non-variety bundle had 12 packs. The packs in each non-variety bundle were identical and came in the most popular flavor based on a pretest (for coffee: Latte flavor; for cookies: chocolate flavor). As the unit price of each pack was controlled to be the same, the non-variety bundle had the same price as the variety bundle in the large-quantity group, while it was one-third the price of the variety bundle in the small-quantity group.

4.1.1. Sample

A total of 143 college students (59.4% male, mean age = 23.8 years) took part in the study for financial compensation.

4.2. Procedure and measures

4.2.1. Procedure

The participants were randomly assigned to one of the five conditions (i.e., four treatment groups and one control group). The participants in the SE condition first read the verbal and pictorial descriptions of the variety bundle and then were asked to evaluate the bundle in terms of favorability and purchase intention. The participants also indicated their expected satiation of consuming the variety bundle. Finally, they were asked to provide some demographic information. In the four JE conditions, the procedures were similar except that the participants were exposed to the descriptions of two stimuli (i.e., the comparison non-variety bundle and the variety bundle) simultaneously. Then, the participants were asked to indicate their evaluations and expected satiation for both the non-variety bundle and the variety bundle. At the end, they were debriefed, compensated, and dismissed.

4.2.2. Measures

We used the same measures for consumer favorability and purchase intention as in study 1. In addition, three items based on 7-point scales were used to measure the participants’ anticipated satiation (Cronbach’s α = .72): how enjoyable they expect it to be when consuming the bundle (1 = I would hate it; 7 = I would love it; reverse coded) (Galak, Kruger, & Loewenstein, 2011); how bored they expect to be after consuming the bundle (1 = not at all bored; 7 = extremely bored); and when they would like to consume the product again (after consuming the bundle) (1 = right now; 7 = not for a while) (Redden & Galak, 2013).

4.3. Results and discussion

We ran a two-way ANOVA for the four treatment groups and found that the participants’ evaluation of the variety bundle was significantly affected by the quantity of repetitive items in the comparison bundle (for favorability: F(1, 112) = 8.91, p < .01; for purchase intention: F(1, 112) = 14.34, p < .001). Specifically, their evaluation of the variety bundle was higher when it was juxtaposed against a bundle with a large quantity (i.e., 12 packs) than against a bundle with a small quantity (i.e., four packs) of items (for favorability: Msmall = 4.79 vs. Mlarge = 5.44; for purchase intention: Msmall = 45.83 vs. Mlarge = 61.76). In addition, the category of the comparison bundle also had a significant impact on the participants’ evaluation of the variety bundle (for favorability: F(1, 112) = 5.29, p < .05; for purchase intention: F(1, 112) = 7.89, p < .01). In particular, the participants gave a higher evaluation of the variety bundle when it was evaluated side-by-side against a non-variety bundle from the same category than that from a different category (for favorability: Msame = 5.36 vs. Mdifferent = 4.87; for purchase intention: Msame = 59.71 vs. Mdifferent = 47.88).
More importantly, the participants’ evaluation of the variety bundle was more susceptible to the quantity of items in the non-variety bundle when the two bundles came from the same category (for favorability: $M_{\text{small}} = 4.85$ vs. $M_{\text{large}} = 5.86$, $F(1, 57) = 12.68$, $p < .001$; for purchase intention: $M_{\text{small}} = 47.22$ vs. $M_{\text{large}} = 72.19$, $F(1, 57) = 27.28$, $p < .001$) than when they came from different categories (for favorability: $M_{\text{small}} = 4.73$ vs. $M_{\text{large}} = 5.00$, $F(1, 55) = .68$, $p > .40$; for purchase intention: $M_{\text{small}} = 44.43$ vs. $M_{\text{large}} = 51.33$, $F(1, 55) = .98$, $p > .30$) (as shown in Fig. 2). The interaction between quantity and category was marginally significant for bundle favorability ($F(1, 112) = 3.07$, $p = .08$) and significant for purchase intention ($F(1, 112) = 4.61$, $p < .05$). This finding supports our conjecture that the variety bundle is less effective in countering anticipated satiation of a non-variety bundle from a different category and is consistent with previous research showing that recalling the variety of items one had in the past accelerates recovery from satiation only when these items belong to the same category as the satiated item (Galak et al., 2009).

Bonferroni post hoc tests for the four treatment groups showed that the participants’ evaluation of the variety bundle in the large-quantity/same-category comparison condition was significantly higher than the evaluations in the other three conditions (for favorability: $p < .01$; for purchase intention: $p < .001$), while the evaluations of the variety bundle in the other three conditions were not significantly different from each other (for both favorability and purchase intention: $p > .10$). This finding indicates that the interaction effect between quantity and category was mainly driven by the higher evaluation of the variety bundle in the large-quantity/same-category comparison condition. In addition, we compared the control group (i.e., SE condition for the variety bundle) with the four treatment groups and found significant differences for two of the treatment groups. In particular, the participants’ evaluation of the variety bundle in the SE mode (for favorability: $M = 4.44$; for purchase intention: $M = 40.81$) was marginally lower than that in the large-quantity/different-category comparison condition (for favorability: $M = 5.00$, $p = .08$; for purchase intention: $M = 51.33$, $p = .08$) and significantly lower than that in the large-quantity/same-category comparison condition (for favorability: $M = 5.88$, $p < .001$; for purchase intention: $M = 72.19$, $p < .001$). This result implies that a large quantity of items in the non-variety bundle drives consumers’ concern for satiation and leads them to evaluate the variety bundle more favorably, particularly when the non-variety bundle is from the same category as the variety bundle.

To further verify the underlying mechanism of concern for satiation, we conducted a mediated moderation test. We expect the interaction between quantity and category on the evaluation of the variety bundle to be mediated by the satiation manifested in the JE mode. We first constructed a satiation measure (i.e., “resultant satiation”) by subtracting the anticipated satiation related to the variety bundle from that activated by the non-variety bundle. The resultant satiation not only changed with the variation of satiation activated by the non-variety bundle, but also reflected the degree of satiation that could be counteracted by the variety bundle. We performed the mediated moderation analysis following the procedure by Hayes (2013) and conducted a mediated moderation test. We expect the interaction effect between quantity and category on the evaluation of the variety bundle to be mediated by the resultant satiation (for favorability: $\beta = .52$, lower 95% CI = .13, upper 95% CI = 1.04; for purchase intention: $\beta = 10.41$, lower 95% CI = 2.23, upper 95% CI = 21.45).

These findings are helpful in understanding how consumer evaluation of the variety bundle in the JE mode is influenced by the comparison non-variety bundle. That is, owing to consumer concern for satiation elicited by the non-variety bundle, the value of variety in countering potential satiation is manifested in the JE mode, leading consumers to favor the variety bundle. Nonetheless, study 2 focused on the JE mode and did not directly examine how the moderating effect of evaluation mode varied with the activated satiation. Furthermore, in study 2, the manipulation of the activated satiation in the non-variety bundles might simultaneously alter other unobserved variables in addition to satiation. Finally, in both study 1 and study 2, the items in the non-variety bundles came in the most popular flavor but might not be the most preferred flavor for each individual participant. We conducted study 3 to further test the mechanism of satiation and address these issues.

5. Study 3: How information associated with repetition influences the effect of evaluation mode

Study 2 showed that consumers’ preference for the variety bundle in the JE mode was driven by their concern for satiation from the non-variety bundle. Study 3 seeks to further examine how the moderating effect of evaluation mode works through the mechanism of satiation, by manipulating the information associated with repetition. In particular, as repetitive consumption of the same items does not necessarily result in satiation (Fishbach et al., 2011), particularly among consumers with strong and consistent preferences, the presence of the non-variety bundle may not always activate the concern for satiation. We posit that if the association between repetitive consumption and satiation is highlighted, the non-variety bundle can easily activate consumer concern for satiation; however, if the association between repetition and consistent preferences (e.g., loyalty) is highlighted, the non-variety bundle will be less likely to activate satiation. That is, associating loyalty rather than satiation with repetition may diminish or even eliminate the effect of evaluation mode on preference for variety. In addition, while the non-variety bundle in study 1 consisted of the most popular flavor identified in a pretest, this flavor may not be the most preferred one for each individual respondent. Therefore, in study 3, we adopt another method to construct.
the non-variety bundle such that it comes in each participant’s personal favorite flavor. We investigate whether consumers’ desire for variety can still override their preference under this circumstance.

5.1. Design and sample

5.1.1. Experimental design

The stimuli in this study were similar to those in study 1 (i.e., potato chips). However, rather than asking the participants to evaluate five given flavors of potato chips, we asked them to indicate their favorite flavor in a filler task before the main experiment. As this experiment was conducted on the computer, the participants’ favorite flavor was automatically entered into the description of the non-variety bundle in the main experiment. The only difference between the two bundles was the presence versus the absence of variety in flavor. Accordingly, this study used a 2 (information associated with repetition: satiation vs. loyalty) × 3 (evaluation mode: SE for the variety bundle vs. SE for the non-variety bundle vs. JE between-subjects design).

5.1.2. Sample

Participants were recruited by an online advertisement on the bulletin board system of a large public university and were compensated for their involvement. The final sample had 206 subjects (43.2% male, mean age = 22.6 years).

5.2. Procedure and measures

The participants were randomly assigned to one of the six conditions. In each condition, they were asked to complete a series of tasks described as independent of each other. All the tasks were completed on the computer. The first task was depicted as a manufacturer’s survey aimed at understanding consumer preferences for potato chips. The participants were asked to indicate their familiarity with and favorability for potato chips as well as their favorability for each flavor listed (which included their personal favorite flavor identified in the preceding filler task). Subsequently, the participants were exposed to the priming task of highlighting different associations with repetition, which was described as an independent “behavior analysis” task. Following the method by Fishbach et al. (2011), the material used in this task was about a college student’s repetitive consumption behaviors, such as being attired in casual sportswear on most days, purchasing the same type of ball pen, and drinking a cup of coffee almost every morning. After reading the material, the participants in the satiation-associated group were asked to rate the extent to which the student was “repetitive,” “boring,” and “dull,” which activated the satiation-based construal for repetitive consumption. However, the participants in the loyalty-associated group were asked to rate the extent to which the student was “loyal,” “strong-minded,” and “sticking to herself/himself,” which activated the preference-based construal for repetitive consumption.

These traits were embedded along other irrelevant descriptions (i.e., “optimistic” and “ambitious”) that were identical in both conditions, and were used to minimize the participants’ awareness of the nature of the priming task. Then, the participants were asked to complete the main experiment, indicating their evaluations of certain bundle(s) of potato chips. The participants in the SE conditions were asked to evaluate either the non-variety bundle that came in their most preferred flavor or the variety bundle that came in five different flavors. The participants in the JE condition were asked to simultaneously evaluate the non-variety and variety bundles. At the end, all participants were debriefed, compensated, and dismissed. The measures for consumer favorability, purchase intention, and concern for satiation for the bundles were the same as those in study 2.

5.3. Results and discussion

5.3.1. Manipulation check

As boringness/satiation is usually viewed negatively, the participants in the satiation-associated condition rated the student lower (M = 4.28) than those in the loyalty-associated condition (M = 4.83; F(1, 204) = 13.00, p < .001). This result suggests that our priming successfully influenced the information contextually associated with repetition, which subsequently biased the participants’ concern for satiation activated by the non-variety bundle. Specifically, the participants’ concern for satiation with the non-variety bundle was significantly higher in the satiation-associated condition (M = 5.49) than in the loyalty-associated condition (M = 3.42; F(1, 178) = 73.55, p < .001).

5.3.2. Results

As expected, we found that the impact of evaluation mode was greater in the satiation-associated condition than in the loyalty-associated condition (for purchase intention: t(200) = 2.29, p < .05), although this effect was nonsignificant for bundle favorability (t(200) = 1.58, p < .10). Specifically, when the association between repetition and satiation was highlighted, we found a significant moderating effect of evaluation mode. In the SE mode, the evaluation of the variety bundle was higher than that of the non-variety bundle (for favorability: M_{variety} = 4.82 vs. M_{non-variety} = 4.31, t(67) = 1.77, p = .08; for purchase intention: M_{variety} = 56.53 vs. M_{non-variety} = 43.28, t(67) = 2.82, p < .01). In the JE mode, the evaluation of the variety bundle was also significantly higher than that of the non-variety bundle (for favorability: M_{variety} = 5.65 vs. M_{non-variety} = 4.16, t(36) = 5.50, p < .001; for purchase intention: M_{variety} = 74.78 vs. M_{non-variety} = 44.00, t(36) = 6.61, p < .001). In particular, compared with the SE mode, the difference between the variety bundle and the non-variety bundle was significantly enhanced in the JE mode (for favorability: t(103) = 2.48, p < .05; for purchase intention: t(103) = 2.65, p < .01), replicating the moderating effect of evaluation mode on preference for variety. Furthermore, the evaluation of the variety bundle was significantly higher in the JE mode than in the SE mode (for favorability: F(1, 69) = 15.60, p < .001; for purchase intention: F(1, 69) = 19.09, p < .001), while the evaluation of the non-variety bundle did not differ across the two modes (for favorability: F(1, 70) = .25, p > .60; for purchase intention: F(1, 70) = .03, p > .80), as shown in fig. 3. This result suggests that the higher evaluation of the variety bundle in the JE mode drives the moderating effect.

When the association between repetition and loyalty was highlighted, the non-variety bundle in the JE mode failed to activate concern for satiation, resulting in no additional preference for the variety bundle. Specifically, the evaluation of the variety bundle was not significantly different from that of the non-variety bundle in the SE mode (for favorability: M_{variety} = 4.88 vs. M_{non-variety} = 4.91, t(62) = .11, p > .90; for purchase intention: M_{variety} = 55.16 vs. M_{non-variety} = 51.63, t(62) = .85, p > .30). There was also no significant difference between the two bundles in the JE mode (for favorability: M_{variety} = 4.94 vs. M_{non-variety} = 4.86, t(35) = .32, p > .70; for purchase intention: M_{variety} = 60.00 vs. M_{non-variety} = 60.50, t(35) = .09, p > .90). Furthermore, the difference between the two bundles in the JE mode was not different from that in the SE mode (for favorability: t(97) = .30, p > .70; for purchase intention: t(97) = .60, p > .50). That is, the moderating effect of the evaluation mode was eliminated when repetition was associated with loyalty rather than satiation.

5.3.3. Discussion

In study 3, where the items in the non-variety bundle came in the favorite flavor for each participant, the moderating effect of evaluation mode was significant only when the repetitive consumption was associated with satiation. This finding is similar to that in study 1. Thus, study 3 reveals that the concern for satiation is innately
inferred from the non-variety bundle in the JE mode. However, when repetition was associated with loyalty, it suppressed the activation of anticipated satiation from the non-variety bundle, and the moderating effect of evaluation mode disappeared. Consistent with study 2, these findings provide additional evidence for our proposition that it is the concern for satiation activated by the non-variety bundle that results in the higher evaluation of the variety bundle in the JE mode than in the SE mode. That is, the moderating effect of evaluation mode on preference for variety works through the mechanism of concern for satiation.

6. General discussion

The present research finds that consumer evaluation of the variety bundle (relative to the non-variety bundle) is significantly higher in the JE mode than in the SE mode, and the higher satiation activated by the comparison non-variety bundle leads to a higher consumer evaluation of the variety bundle in the JE mode. In essence, the moderating effect of evaluation mode on consumer preference for variety is driven by the mechanism of concern for satiation. These findings are new and contribute to the literature on both consumer variety-seeking behavior and the marketing of product bundles.

Given the prevalence of product bundles in supermarkets and other retail outlets, our findings have considerable implications for marketers. First, our research suggests that retailers will benefit by incorporating variety in the product bundles that they carry. To counteract anticipated satiation during consumption, consumers generally favor product bundles with variety. Assuming that incorporating variety in product bundles incurs little additional cost (relative to the non-variety bundle), it makes sense for retailers to incorporate different formats of variety (e.g., flavor, scent, and shape) in their bundles.

More importantly, variety has to be complemented with proper display strategies for consumers to fully appreciate its value. We find that consumers better appreciate the value of the variety bundle when it is displayed alongside the non-variety bundle than when it is displayed by itself. This is due to the satiation activated by the non-variety bundle in the JE mode, which enhances consumer evaluation of the variety bundle. Thus, retailers seeking to drive more sales may wish to juxtapose variety bundles against non-variety bundles on store shelves. To illustrate, the value of a variety yogurt bundle would be better appreciated when it is placed next to a non-variety yogurt bundle. This effect is particularly salient if the non-variety bundle contains a large quantity of items and is from the same product category as the variety bundle.

In addition to the joint display strategy, retailers can also enhance the relative advantage of the variety bundles to the non-variety bundles by associating repetitive consumption with satiation. To illustrate, for perishable goods such as snacks, retailers can frame the expiration date to appear closer, such that consumers will prefer the variety bundle more. In addition, firms may also use advertising at the retail outlet or in-store slogans to activate consumers’ satiation-based construal of consumption (Fishbach et al., 2011) and, thus, increase their preference for the variety bundles.

References


Does private-label production by national-brand manufacturers create discounter goodwill?

Anne ter Braak a,⁎, Barbara Deleersnyder b,1, Inge Geyskens b,2, Marnik G. Dekimpe b,a,3

a University of Leuven, Naamsestraat 69, 3000 Leuven, Belgium
b Tilburg University, Wurandel 2, 5000 LE Tilburg, the Netherlands

A R T I C L E   I N F O
Article history:
First received in 21, June 2012 and was under review for 5 months
Available online 7 July 2013
Area Editor: Peter J. Danaher

Keywords:
Private label
Shelf allocation
Sourcing decisions
Discounters
Channel relationships

A B S T R A C T
Discount stores have a private-label dominated assortment where national brands have only limited shelf access. These limited spots are in high demand by national-brand manufacturers. We examine whether private-label production by leading national-brand manufacturers for two important discounters (one hard and one soft) creates discounter goodwill. We estimate a selection model that is based on a sample of 450 manufacturer-category combinations from two leading discounters (Aldi in Germany and Mercadona in Spain), and we show that private-label production is indeed rewarded: national-brand manufacturers that are involved in such practices have a higher likelihood of procuring shelf presence for their brands. Moreover, while powerful manufacturers are intrinsically more likely to obtain shelf presence with soft discounters, manufacturers with less power can compensate for this by producing private labels. No such dependence on power exists for hard discounters. However, not all national-brand manufacturers are equally likely to produce private labels for discounters. We find that national-brand manufacturers are less likely to do so when: (a) they experience more sales growth, (b) it is more difficult to produce high-quality products in a specific category, (c) they invest more advertising support into their brands, and (d) they introduce more innovations. Moreover, a higher price differential relative to the discounter’s private labels makes national-brand manufacturers less likely to engage in private-label production for hard discounters.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Private labels are becoming increasingly important in grocery retailing. They already account for 43% of the total consumer packaged goods (CPG) consumption in the U.K., 32% in Germany, 31% in Spain, and 17% in the U.S. (ACNielsen, 2011). Absolute sales numbers are equally impressive. Wal-Mart’s U.S. private-label sales, for example, were expected to generate approximately $90 billion in 2012 (PlanetRetail, 2012a). In 2011, the value of the European private-label food market reached €436 billion (PlanetRetail, 2011).

Private-label production opportunities have arisen in response to these large volumes. A substantial part of this production is accounted for by dedicated private-label manufacturers. However, several national-brand manufacturers – such as Alcoa (which owns Reynolds Wrap aluminum foil), Parmalat, and H.J. Heinz – have adopted an “if you can’t beat them, join them” attitude, as they now engage in private-label production (Dunne & Narasimhan, 1999; Quelch & Harding, 1996). In the U.S., it has been estimated that over half of the national-brand manufacturers pursue a dual strategy, engaging in private-label production in addition to their national-brand activities (Kumar & Steenkamp, 2007).

A major motivation for national-brand manufacturers to engage in private-label production is, as the former CEO of Ontario Foods testified, “to cultivate a better relation with retailers” (Littman, 1992, p. 2). According to Dunne and Narasimhan (1999), private-label production represents a neglected opportunity for manufacturers that seek closer ties with their retailers, offering a chance to create retailer goodwill. IGD’s European Private Label Survey revealed that 47% of all of the surveyed suppliers believe that strengthening their relationship with the retailer could be a major advantage of supplying private labels (IGD, 2006). However, there is substantial variability in the willingness of brand manufacturers to produce private labels. While some manufacturers, such as Dole and Kraft, are eager to do so, other manufacturers, such as Coca-Cola and Heineken, have explicitly stated that they will never engage in private-label production. This variation in willingness may be driven by differences in the manufacturers’ sales growth (Kumar & Steenkamp, 2007), in their ability to...
influence the produced private-label quality (e.g., by lowering private-label quality or by producing a private label that imitates a competitive national brand to preserve their own sales) (Dunne, 1999), or in the level of brand image that is at stake for them (de Jong, 2007). Moreover, a retailer may be more inclined to collaborate with some national-brand manufacturers to produce its private labels than with others, as some may provide more quality assurance (Sethuraman, 2009) or have higher innovative capacities (Kumar & Steenkamp, 2007) than others. In other instances, a retailer may prefer to assign its private-label production to a dedicated producer (that does not have any national brands of its own), rather than to a dual brander, as the former are known to be more cost-focused. Whether a given national-brand manufacturer produces private labels for a given retailer is therefore driven by manufacturer as well as retailer considerations. This issue will be reflected in our subsequent theorizing.

In this study, we look at the antecedents and consequences of private-label production by national-brand manufacturers for retailers. This topic has recently been recognized as being of great managerial interest but seriously lacking in empirical research (Sayman & Raju, 2007, p. 147; Sethuraman, 2009, p. 771; Sethuraman & Raju, 2012, p. 331). The lack of empirical insights can be attributed to the secrecy surrounding the question of which manufacturer produces retailers’ private labels, which makes it notoriously difficult to obtain the required data (Sethuraman & Raju, 2012). We undertook a massive data collection effort, covering detailed information on over 400 manufacturer-category combinations and two retailers, to empirically test whether private-label production by national-brand manufacturers indeed leads to retailer goodwill, while also considering the potential drivers of private-label production.

We test our hypotheses in a discount setting, where we focus on the prototypical hard discounter, Aldi, and on Europe’s largest soft discounter, Mercadona. Discounters are characterized by a limited assortment that is dominated by private labels, relatively small shopping areas, and very competitive prices.4 To offer lower prices, they use a simplified ‘no-frills’ store format. Hard discounters typically offer fewer than 1400 SKUs in stores of approximately 1000 square meters. Soft discounters, in contrast, have a more extended range of meters. Soft discounters, in contrast, have a more extended range of

4 As such, they are distinct from large-scale, every-day-low-price (EDLP) retailers such as Wal-Mart in the U.S. and from large supermarkets as Carrefour and Tesco in Europe (Cleeren et al., 2010).

5 Aldi relied exclusively on its private labels for many years (Brandes, 2005).

Second, the discount setting allows us to test our ideas in a more controlled, natural experiment-like setting. While many alternative manifestations of retailer goodwill (in addition to shelf presence) should be taken into account at traditional retailers such as reduced slotting allowances, lower promotional fees, a higher extent of pass-through, and/or more promotional feature and display support (most of which also constitute data that are difficult to obtain), this is not the case for discounters. Discounters typically do not charge slotting allowances or promotional fees, and their format is characterized by very limited promotional activity (PlanetRetail, 2010). Because one form of retailer goodwill (e.g., shelf presence) is especially plausible for brand manufacturers that work with discounters, the discount setting offers a cleaner and more controlled environment for studying the effects of private-label production by national-brand manufacturers on retailer goodwill.

We develop a model that allows us to test whether brand manufacturers that are involved in private-label production for a discounter have a higher likelihood of procuring shelf presence for their national brands. The paper is organized as follows. First, we review the limited literature on private-label production by brand manufacturers. Then, we present our conceptual framework and hypotheses. Next, we discuss the research method, data, and empirical findings. The final section discusses the implications for researchers and managers and provides suggestions for further research.

2. Literature review

Despite “the richness of the phenomenon and the high level of managerial interest,” surprisingly little research has studied private-label (PL) production by national-brand (NB) manufacturers (Sethuraman, 2009, p. 771). An initial set of papers has developed game-theory models to study the issue. Kumar, Radhakrishnan, and Rao (2010), for example, consider the retailer’s decision to work with either a dedicated PL supplier or a dual brander. They focus on a two-level supply chain including a retailer, a NB manufacturer, and a dedicated (independent) manufacturer. Depending on whether the dedicated supplier or the NB manufacturer produces the retailer’s PL products, a two-vendor or one-vendor regime emerges. Based on the size of the retailer’s quality-versus price-sensitive customer segment, the retailer is shown to be better off with either the NB manufacturer or a dedicated PL manufacturer. Retailers with a large price-sensitive customer segment would not prefer a NB manufacturer to supply their PL products: NB manufacturers would lower PL quality (to preserve their own sales), and this is only profitable for the retailer if the quality-sensitive (NB-prone) customer segment is large enough. Interestingly, Kumar et al. (2010, p. 156) note that their analysis does not apply to discount stores, as these are very focused on the low-end, highly price-sensitive, customer segment.

Other studies have approached the PL-production decision from the point of view of a manufacturer that has to decide whether to engage in PL production for a given retailer. Using a model of vertical differentiation, Gomez-Arias and Bello-Acebron (2008) derive that, depending on the PL’s quality positioning, a NB manufacturer might be more or less willing to supply the PL product. Gomez-Arias and Bello-Acebron distinguish between high-quality and low-quality NB manufacturers. The pressure to produce PLs is highest when the PL enters the market at a quality level that is similar to the manufacturer’s NB quality. A high-quality NB manufacturer then wants to produce a high-quality PL to pre-empt competition and to be able to position the PL such that it competes with other NBs and not its own. Wu and Wang (2005), in turn, show analytically how PL production by one NB manufacturer can be used to mitigate promotional competition with other NB manufacturers. However, this study may be less applicable in a typical EDLP-based discount setting in which promotions are largely absent to start with.
Moreover, none of the aforementioned studies provides empirical support for their various contentions. This is not surprising, given that most manufacturers keep their involvement in PL production confidential (de Jong, 2007) out of fear for the impact that such knowledge may have on their main brands (Gomez-Arias & Bello-Acebron, 2008) and on their other retail relationships. This situation makes empirical research on the topic difficult, and it has prompted Sethuraman (2009, pp. 771–773) to call for more empirical research on both the antecedents and the consequences of dual branding (i.e., the practice of NB manufacturers to also engage in PL production). More recently, Sethuraman and Raju (2012, p. 351) again concluded that “we need a better understanding of why a manufacturer would supply private labels and why a retailer would accept the same.” In line with these calls, Chen, Narasimhan, John, and Dhar (2010) have estimated a structural model to derive the profit implications for various PL supply arrangements, and they have shown through a number of policy simulations, in the context of a single market (fluid milk), how engagement in PL production may be beneficial to NB manufacturers.

Our analysis differs from the above-referenced studies in a number of ways. First, while previous studies (e.g., Chen et al., 2010; Gomez-Arias & Bello-Acebron, 2008; Kumar et al., 2010) considered the (expected) profit consequences from various PL-supply schemes, we focus on the potential goodwill creation of a PL-production decision, which has been recognized as a key strategic consideration in both the academic literature (Dunne & Narasimhan, 1999) and in the business press (IGD, 2006). Second, we empirically examine why certain NB manufacturers are more likely to be involved in the production of a discounter’s PLs than their competitors, and we pursue this question across a broad set of grocery categories and manufacturers. We study this problem at two leading European discounters. To the best of our knowledge, no such large-scale empirical research is available on the determinants of a manufacturer’s engagement in PL production. Finally, several of the aforementioned analytical models (e.g., Kumar et al., 2010; Wu & Wang, 2005) are less suited for use in a (EDLP-based) discount context. We focus on discounters, and we therefore contribute to the (thus far) limited empirical literature (Cleeren, Verboven, Dekimpe, & Gielens, 2010; Deleersnyder et al., 2007) on this fast-growing retail format. As such, our study also differs substantially from ter Braak, Dekimpe, and Geyskens (2013), who study the implications of dual branding on retail profit margins in the context of a conventional supermarket, which follows a very different business model than discount stores.

3. Conceptual framework and hypotheses

Fig. 1 summarizes our theorizing, both on the relationship between PL production and shelf presence, and on the determinants of a NB manufacturer’s participation in such production.

3.1. The effect of PL production on shelf presence

PL production by a NB manufacturer may be seen by the discounter as a pledge by the NB manufacturer. “Pledges” are actions that are undertaken by channel members that demonstrate good faith and that bind channel members to the relationship (Anderson & Weitz, 1992, p. 20). The offering of pledges functions as a signal of goodwill and invites reciprocal actions. By making the pledge of PL production, a NB manufacturer effectively signals its cooperative behavior (Dunne & Narasimhan, 1999). This may lead to retailer goodwill, which can originate during the negotiation process and/or can become more pronounced as the relationship develops. This is consistent with the reciprocity literature (see, e.g., Uhl-Bien & Maslyn, 2003, p. 514), which has argued that the time span of reciprocation can range from high immediacy (i.e., instantaneous) to low immediacy. Following this logic, PL production may represent a good opportunity for NB manufacturers to enhance the likelihood of obtaining access to the discounter’s most valued, yet scarce, resource, i.e., shelf presence. Hence:

H1. PL production by a NB manufacturer increases the likelihood of shelf presence.

3.2. The effect of NB manufacturer market power on shelf presence

When testing H1, we control for the intrinsic market power of the NB manufacturer. Following Shervani, Frazier, and Challagalla (2007), we define manufacturer market power as a manufacturer’s ability to influence the actions of others in a product market. The industrial-organization literature has argued that a firm’s market power is based on its market position, as reflected by its market share. If a firm operates...
in multiple product markets, its market power can vary considerably across them. Empirical research has shown that powerful firms are able to secure relatively high levels of influence on the behavior of related channel members (e.g., Anderson, Lodish, & Weitz, 1987; Shervani et al., 2007; Sudhir & Rao, 2006). In a similar vein, we argue that more powerful manufacturers – with a larger volume share in the category that is obtained from more and stronger brands – are generally in a better position to secure shelf space. The addition of their NB is expected to improve a discounter’s perceived assortment quality and variety, as these brands will stand out more against an otherwise PL-dominated assortment (Deeleersnyder et al., 2007). Because of their expected contribution to the discounter’s financial goals, discounters will be more receptive to brands that are offered by powerful NB manufacturers. We therefore postulate:

**H2.** NB manufacturer market power increases the likelihood of shelf presence.

Moreover, this variable may interact with PL production. The positive effect of PL production on the likelihood of shelf presence is expected to be more pronounced for less powerful NB manufacturers. Relationship quality has been shown to have a larger impact on new-product acceptance when products are moderately attractive than when they are very attractive (as when they are offered by the most successful manufacturer) (Kauffman, Jayachandran, & Rose, 2006). In a similar vein, retailers have been found to prefer the leading NB manufacturer as category captain unless smaller players can (or are willing to) offer a special service (Subramanian, Raju, Dhar, & Wang, 2010). As such, PL production can be used by smaller NB manufacturers as a tool to overcome an inherent market-power disadvantage. In other words, when a NB manufacturer’s market power is lower, the likelihood that a discounter will grant shelf presence to that NB manufacturer increases considerably when the manufacturer establishes a strong relationship with the discounter (i.e., it produces PL products for the discounter). When its market power is very high, a NB manufacturer is already likely to be granted shelf presence, regardless of PL production. Therefore, the latter has less incremental impact. Following this line of reasoning:

**H3.** NB manufacturer market power diminishes the effect of PL production in increasing the likelihood of shelf presence for the NB manufacturer.

### 3.3. Drivers of PL production

We include three major drivers of PL production by a NB manufacturer for a discounter, reflecting three categories of motives: the extent of sales growth (economic motive), the relative ease of producing high-quality products in the category (quality/positioning motive), and the marketing tools that are used to support the manufacturer’s brands (brand-equity motive).

#### 3.3.1. Manufacturer sales growth

PL production often starts on an opportunistic basis when a manufacturer has idle capacity (Gomez-Arias & Bello-Acebron, 2008; Kumar & Steenkamp, 2007), which may, for example, be caused by increased competition from other manufacturers or by reduced demand during tough economic times (Sethuraman & Raju, 2012). A PL order is then used to complement the lower NB sales. PL production can increase or maintain NB manufacturers’ total sales volumes and thereby safeguard profits (Quelch & Harding, 1996; Soberman & Parker, 2006).

For a discounter, we could argue that (all else being equal) it may also prefer to work with a manufacturer with lower NB sales growth. Indeed, such manufacturers could be perceived as easier targets through which to achieve good supply conditions, thereby increasing the discounter’s profitability.

We therefore hypothesize that higher sales growth decreases a manufacturer’s likelihood of engaging in PL production. Negative NB sales growth, in contrast, may lead to costly excess capacity for a manufacturer and hence increase a manufacturer’s likelihood of producing PLs.

**H4.** Manufacturer sales growth decreases the likelihood of observing PL production by the manufacturer for the discounter.

#### 3.3.2. Ease of producing high-quality products

A second key consideration for NB manufacturers is the possibility of managing and influencing PL quality (Dunne, 1999; Kumar & Steenkamp, 2007). When it is difficult to produce high-quality products in a category, discounters will have a harder time matching the intrinsic quality of the NBs, which gives the NBs a competitive advantage. To preserve that advantage, NB manufacturers will be reluctant to enter into a PL-production relationship with a discounter. By contrast, when it is relatively easy for discounters to match NB quality, NB manufacturers will be more willing to engage in PL production, as there is no sustainable competitive advantage at stake. Gomez-Arias and Bello-Acebron (2008) showed that a NB manufacturer’s incentive to produce PLs is particularly high if the PL matches a manufacturer’s NB quality, which is more likely when the ease of producing high-quality products is high. NB manufacturers then have the ability to influence the positioning of PLs (Gomez-Arias & Bello-Acebron, 2008; Kumar et al., 2010), for example by producing a PL that imitates a competitive NB to preserve their own sales (Sayman, Hoch, & Raju, 2002). Moreover, if they do not produce a discounter’s PL, a competitor is likely to do so, which gives NB manufacturers an incentive to pre-empt the competition (Steenkamp & Dekimpe, 1997).

However, to come to actual PL production for a given discounter, both the manufacturer and the discounter must agree. Indeed, the latter must decide whether to work with a dedicated PL manufacturer (which does not sell any NBs) or to select a dual brander. Dual branders are said to have more innovative capacity (Kumar & Steenkamp, 2007) and to offer more quality assurance (Sethuraman, 2009) than dedicated manufacturers. Clearly, the second issue is especially relevant when it is difficult to produce high-quality products. Hence, when it is difficult to produce high-quality products, a discounter will more likely opt for a leading NB manufacturer for its PL production. In contrast, when production is relatively easy, a discounter may well opt for a dedicated manufacturer that is more price-focused (Kumar & Steenkamp, 2007).

Depending on the relative strength of both arguments, the observed choice will be driven more by manufacturer (positive association between ease of producing and PL production) or discounter (negative association between ease of producing and PL production) considerations. If the effect is positive, the manufacturer considerations dominate; if it is negative, the discounter rationale prevails. In sum:

**H5a/b.** The higher the ease of producing high-quality products in the category, the higher/lower the likelihood of observing PL production by a NB manufacturer for the discounter.

#### 3.3.3. Manufacturer marketing tools

Numerous studies discuss how NB manufacturers can successfully compete against PLs. According to Kumar and Steenkamp (2007), there are three distinct tools that are used by NB manufacturers to create winning value propositions for their NBs. First, NB manufacturers may try to differentiate themselves from cheaper PL imitations by conveying unique emotional benefits and signaling future demand through advertising (Desai, 2000; Steenkamp & Dekimpe, 1997). However, this carefully built-up image may be damaged if their
customer base discovers that they also produce PLs (Gomez & Benito, 2008; Hoch, 1996), making NB manufacturers that rely heavily on advertising, reluctant to engage in PL production. Second, charging a hefty price premium in the market over PLs is another way to position the NBs away from PLs (Ailawadi, Lehmann, & Neslin, 2003). Manufacturers, again, may not be willing to risk this competitive advantage by producing a cheaper variant for the discounter. Finally, they can introduce innovations and thereby offer products that are distinct from any existing PL. Maintaining this quality edge curbs PL growth, as it puts retailers in the position of imitating yesterday’s favorites (Lamey, Deleersnyder, Steenkamp, & Dekimpe, 2012). NB manufacturers that are heavily involved in developing new products will be less inclined to produce PLs, as it may put them in a position where the discounter can exert pressure to share the latest technologies (Dunne & Narasimhan, 1999), which would undermine their competitive advantage. Each of the marketing tools is therefore predicted to decrease the likelihood of PL production by the NB manufacturer for the discounter. As indicated above, some manufacturers motivate their willingness to participate in the PL production process through an “if you can’t beat them, join them” reasoning (Sethuraman & Raju, 2012). The three above-stated arguments reflect this idea in that they describe the conditions under which NB manufacturers are more/less able to withstand the impact of PL growth on their own performance.

Given the secrecy surrounding the identity of the actual PL producers, discounters should not be concerned about the amount of advertising support that is given by the NB manufacturer to its own brands nor about the price differential between the manufacturer's NBs and the discounter’s PLs when deciding on their PL sourcing. However, given that dual branders are thought to provide more innovative capacity when producing PLs (Kumar & Steenkamp, 2007), the innovativeness of a NB manufacturer may play an important part in the discounter’s preference (i) for a dedicated supplier versus a dual brander and (ii) if the latter option is selected, its preference for some NB manufacturers over others. The more innovative a manufacturer is, the higher a discounter’s interest in that manufacturer as the preferred option to produce its PLs. In sum:

**H6.** The more a NB manufacturer advertises, the lower the likelihood of PL production by the manufacturer for the discounter.

**H7.** The higher the price premium a NB manufacturer charges over a discounter’s PLs, the lower the likelihood of PL production by the manufacturer for the discounter.

**H8a/b.** The more innovative a NB manufacturer is, the lower/higher the likelihood of observing PL production by the manufacturer for the discounter.

For hypothesis H8a/b, depending on the relative strength of both arguments, the observed choice will be driven more by manufacturer (negative association between a manufacturer’s innovativeness and PL production) or discounter (positive association between a manufacturer’s innovativeness and PL production) considerations.

### 3.3.4. Manufacturer market power

Finally, we also consider the effect of manufacturer market power on the PL-production decision. We propose that manufacturer market power may have an impact on a NB manufacturer’s likelihood of producing PLs. According to Dunne and Narasimhan (1999), producing PLs is especially interesting for small manufacturers of non-leading NBs that seek to increase their sales volume. Alternatively, one could argue that PLs have become so successful that NB manufacturers that want to produce PLs for discounters need sufficient capacity (size) to handle the large volumes (IGD, 2005). Because good arguments are available to support an increased likelihood of PL production for lower/higher manufacturer market power, we offer competing hypotheses:

**H9a/b.** NB manufacturer market power decreases/increases the likelihood of observing PL production by the manufacturer for the discounter.

### 4. Research setting and measures

#### 4.1. Setting

Consumer packaged goods (CPG) companies regard Germany and Spain as two key European markets with respect to discounters. They are both among the largest consumer markets in Western Europe. Not only does the discount format originate from Germany, its current market share in Germany already exceeds 40% (PlanetRetail, 2012b). Further, discounter share is rapidly increasing in Spain, where it sums up to close to 30% (PlanetRetail, 2012c).

In Germany, we study Aldi, the “mother of all discounters.” Aldi’s market share in Germany was approximately 15% in 2012 (PlanetRetail, 2012d). Aldi operates a total of over 8000 stores across Europe, in 16 countries. With current global sales of $80 billion, Aldi is ranked among the top ten grocery retailers in the world. Although Aldi’s PL assortment differs across countries, it follows a centralized, national approach to its assortment management, with the same offering present in each outlet within a country (ACNielsen, 2007). Traditionally, Aldi did not carry any NBs in its assortment. However, it recently started accepting NBs on a limited basis. Although more than 90% of Aldi’s range still consists of PLs, recent NB additions in Germany include Snickers candy bars (Masterfoods), Del Monte canned fruit (Del Monte), and Quality Street sweets (Nestlé).

In Spain, we study the leading discounter Mercadona. Mercadona has been the most successful store in the country in recent years. It is the largest grocery retailer in Spain, operating approximately 1400 outlets, and it was ranked the ninth most reputable company in the world in 2009 by Forbes Magazine. The firm’s strategy is built around consistently offering value for money rather than short-term price promotions. Its sales in Spain have almost doubled over the last five years, amounting to $24 billion (PlanetRetail, 2012c). Currently, Mercadona sells approximately 2000 PL products. Shelf space for NBs is limited, as Mercadona tries to achieve a 50–50% mix between PLs and popular brands (Bain & Company, 2008), with typically just one or two NBs facing each PL (Fernandez Nogales & Gomez Suarez, 2005). As a result, 30% fewer product varieties tend to be available at Mercadona than can be found in traditional supermarkets (Bain & Company, 2008). While Mercadona can be classified as a soft discounter, Aldi is a prototypical hard discounter.

Clearly, NB shelf space is a scarce resource for both discounters, which is sought after by many NB manufacturers. Interestingly, discounters typically do not charge slotting allowances or promotional fees (PlanetRetail, 2010). Hence, these issues do not enter channel (goodwill) negotiations. By considering both Europe’s leading hard (Aldi) and soft (Mercadona) discounter, we can infer to what extent our findings differ (or are common) across both format types.

#### 4.2. Sample and measures

Our unit of analysis is the NB manufacturer in a certain category at the discounter. We only consider categories that contained at least one NB and thus, where the discounter made the deliberate decision to extend its assortment with a NB offering. For Aldi, we collected data on 37 grocery categories. For each category, we determined the discounter’s PL assortment across countries, it follows a centralized, national approach to its assortment management, with the same offering present in each outlet within a country (ACNielsen, 2007). Traditionally, Aldi did not carry any NBs in its assortment. However, it recently started accepting NBs on a limited basis. Although more than 90% of Aldi’s range still consists of PLs, recent NB additions in Germany include Snickers candy bars (Masterfoods), Del Monte canned fruit (Del Monte), and Quality Street sweets (Nestlé).

In Spain, we study the leading discounter Mercadona. Mercadona has been the most successful store in the country in recent years. It is the largest grocery retailer in Spain, operating approximately 1400 outlets, and it was ranked the ninth most reputable company in the world in 2009 by Forbes Magazine. The firm’s strategy is built around consistently offering value for money rather than short-term price promotions. Its sales in Spain have almost doubled over the last five years, amounting to $24 billion (PlanetRetail, 2012c). Currently, Mercadona sells approximately 2000 PL products. Shelf space for NBs is limited, as Mercadona tries to achieve a 50–50% mix between PLs and popular brands (Bain & Company, 2008), with typically just one or two NBs facing each PL (Fernandez Nogales & Gomez Suarez, 2005). As a result, 30% fewer product varieties tend to be available at Mercadona than can be found in traditional supermarkets (Bain & Company, 2008). While Mercadona can be classified as a soft discounter, Aldi is a prototypical hard discounter.

Clearly, NB shelf space is a scarce resource for both discounters, which is sought after by many NB manufacturers. Interestingly, discounters typically do not charge slotting allowances or promotional fees (PlanetRetail, 2010). Hence, these issues do not enter channel (goodwill) negotiations. By considering both Europe’s leading hard (Aldi) and soft (Mercadona) discounter, we can infer to what extent our findings differ (or are common) across both format types.

---

7 This feature is important when collecting data through store visits (cf. infra).
obtained data on 53 categories in which at least one NB was available from Kantar Worldpanel. We retained the top-five NB manufacturers in terms of their 2009 category sales volume in Spain. 8 To illustrate the range of products that are available in our dataset, we grouped the categories into broader product groups. Table 1 shows these groups, along with some illustrative examples of the included categories. For each of the resulting 450 manufacturer–category combinations (the top-five NB manufacturers for 90 categories, with 37 × 5 observations for Aldi and 53 × 5 observations for Mercadona), we combined a wide variety of data sources to operationalize our variables.

4.2.1. Shelf presence

We measure shelf presence on the basis of GfK/Kantar consumer panel data for Germany and Spain, which cover all of the grocery purchases that were made by a representative national sample of 20,000+ German and 8000+ Spanish households. Shelf presence captures whether a NB that was owned by one of the top-five NB manufacturers in a category was available in that category at a given discounter (1 = yes, 0 = no). For Aldi, this information was available between January of 2002 and June of 2008. For the 185 (37 × 5) manufacturer–category combinations, we checked whether the NB manufacturers own a NB for which a purchase record in the category at Aldi could be found during the time frame of our data. We observed shelf presence for 46 (out of 185) manufacturer–category combinations. For Mercadona, shelf presence was measured in the year 2009. In this case, shelf presence was observed for 139 (out of 265) manufacturer–category combinations. Proportionally, more instances of shelf presence for the top-five NB manufacturers are seen with Mercadona, which is consistent with its positioning as a soft discounter. Table 2 provides more details on the distribution of the number of top-five NB manufacturers with shelf presence at the discounter across categories.

4.2.2. PL production

We measure “PL production” as a binary variable that is coded 1 if the NB manufacturer engaged in PL production for the discounter in the category and 0 otherwise. In spite of the secrecy surrounding Aldi (Brandes, 2005) and the reluctance to acknowledge PL production by NB manufacturers (de Jong, 2007), we were able to obtain information on PL production for Aldi through extensive field research. Between January of 2002 and June of 2008, approximately 650 PLs were sold in the 37 categories that were under investigation at Aldi. For each of those 650 PLs, we determined the producer. Because these data were not readily available through conventional channels, various sources were consulted. As a starting point, we obtained four books with information on the manufacturers of 200 popular PLs that are sold at Aldi (Bertram, 2006; Schaab & Eschenbek, 2008; Schneider, 2005, 2006). For the remaining 400 + products, we replicated the procedure described in these books to uncover the manufacturers of all of the PLs in the 37 categories that were studied. To that extent, we exploited the fact that Aldi is one of the rare retailers that prints the address of the manufacturer on its packages (de Jong, 2007). To extend and/or validate this information, we also consulted various websites that are devoted to consumer product reviews, company profiles, and/or discounter products (including www.ciao.de, www.yopi.de, www.discount-archiv.de, and www.wer-zu-wem.de). Specifically, for each PL, we recorded (through numerous store visits and online searches) the manufacturer’s address. We then checked whether the manufacturer’s address matched the address of one of the top-five NB manufacturers in the category. In case of no address match, we looked up the name of the manufacturer that was located at the stated address and assessed whether it was a subsidiary of one of the top-five NB manufacturers. In this way, we account for situations in which NB manufacturers may produce PLs for a discounter at a different address or location and/or at a subsidiary with a different firm name. For 68 (out of 185) manufacturer–category combinations, we found evidence of PL production for Aldi (see Table 3, panel A). For 63 (93%) of those manufacturer–category combinations, more than one data source was found confirming PL production activity. In 32 (or 86%) of the 37 categories that were examined at Aldi, at least one of the top-five NB manufacturers was involved in the production of Aldi’s PLs. 9 Examples for Aldi include Campina (yogurt), Dr. Oetker (dessert), and Bonduelle (canned vegetables).

For Mercadona, we were fortunate to obtain, with the help of Kantar Worldpanel, internal category-level PL producer information for the year 2009. In this case, we again implemented an extensive online search to detect any less obvious dependence between the listed producer and the 53 × 5 NB manufacturers in our sample. For 24 (out of 265) manufacturer–category combinations, we found evidence of PL production for Mercadona (see Table 3, panel B). In 23 (or 43%) of the 53 categories that were examined at Mercadona, at least one of the top-five NB manufacturers was involved in the production of Mercadona’s PLs. 10 For reasons of confidentiality, no Mercadona examples can be revealed. A simple bivariate χ² test provides initial evidence of a relationship between PL production and shelf presence ($\chi^2(1) = 21.33, p < .01$ for Aldi and $\chi^2(1) = 16.27$.

---

8 For Aldi, PL production data covered the 2002 to 2008 period (cf. infra) — we determined the top-five NB manufacturers for the midpoint of this time frame (2005). Because PL production data for Mercadona pertained to 2009, we determined the top-five NB manufacturers for 2009. No PL production data on preceding years were available for Mercadona.

9 In the other five categories, where none of the top-five players produced any of the discounter’s PLs, we found evidence of PL production by a non-top-five NB manufacturer.

10 In 12 out of the 30 remaining categories, Mercadona chose a non-top-five NB manufacturer to produce its PLs.

---

### Table 1

<table>
<thead>
<tr>
<th>Product group</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assorted foods</td>
<td>Noodles, jam, rice</td>
</tr>
<tr>
<td>Beverages</td>
<td>Beer, mineral water, juice</td>
</tr>
<tr>
<td>Candy</td>
<td>Candy bars, chocolate, bonbons</td>
</tr>
<tr>
<td>Canned/bottled foods</td>
<td>Fish, beans, tomato</td>
</tr>
<tr>
<td>Care products</td>
<td>Shampoo, diapers, toilet tissue</td>
</tr>
<tr>
<td>Cleaning products</td>
<td>Bleach, detergent, fabric conditioner</td>
</tr>
<tr>
<td>Cooking fats</td>
<td>Butter, olive oil, margarine</td>
</tr>
<tr>
<td>Dairy products</td>
<td>Milk, yogurt, ice cream</td>
</tr>
<tr>
<td>Household supplies</td>
<td>Basket bin bags, toilet tablets, celluloses</td>
</tr>
<tr>
<td>Instant meals</td>
<td>Ready desserts, ready meals, salad</td>
</tr>
<tr>
<td>Pastry</td>
<td>Cakes, sweet biscuits</td>
</tr>
<tr>
<td>Pet products</td>
<td>Wet dog food, dry dog food</td>
</tr>
<tr>
<td>Taste enhancers</td>
<td>Ketchup, salad dressing, mayonnaise</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Number of top-5 NB manufacturers with shelf presence in the category</th>
<th>% of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative examples</td>
<td>% of categories</td>
</tr>
<tr>
<td>0</td>
<td>38%</td>
</tr>
<tr>
<td>1</td>
<td>27%</td>
</tr>
<tr>
<td>2</td>
<td>14%</td>
</tr>
<tr>
<td>3</td>
<td>16%</td>
</tr>
<tr>
<td>4</td>
<td>5%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
</tr>
</tbody>
</table>

Note: n.a. = not applicable.
4.3. Predictors

We measure the covariates for Aldi in the year preceding shelf presence. When no brand of the NB manufacturer was listed at Aldi, we use 2005 as the base year for the covariates. For Mercadona, we measure the covariates in the year 2008 (i.e., the year before we observe shelf presence). For a complete overview of the timing, source, and operationalization of our variables, we refer to Table 4.

4.3.1. Manufacturer sales growth

Manufacturer sales growth is measured as the maximum growth in a manufacturer’s NB volume sales across three consecutive years in the category and obtained from the panel data. Whereas positive sales growth signals sufficient NB demand and perhaps the need for investment in additional production capacity, negative sales growth signals idle production capacity (Lieberman, 1987). We also specify sales growth as the average growth in a manufacturer’s NB volume sales in the category across the three consecutive years. Our results remain the same.

4.3.2. Ease of producing high-quality products

We measure the perceived ease of producing high-quality products through a 5-point reverse-scored survey item: “In the category XXX, making good quality products is difficult.” We combined this country-specific information (i.e., German consumers for the Aldi observations and Spanish consumers for the Mercadona observations) from regular users of the category (each consumer could provide information on at most four categories), with information from four industry experts (who had to rate all of the categories). Both groups had similar assessments, as evidenced in a highly significant correlation \( r = .41, p < .01 \) between their respective averages.

4.3.3. Manufacturer marketing tools

We measure the extent of “advertising” by means of the annual national advertising expenditures (in €) in the category by the NB manufacturer. The German advertising data were obtained from Thomson Media Control, a provider of market data on advertising spending across all media in Germany. The Spanish advertising data were obtained from InfoAdex, a company that measures advertising investments across various media in Spain. Second, a manufacturer’s “price premium over PLs” was obtained from the panel data and measured as the market-share-weighted average national unit price of all of a manufacturer’s NBs in the category compared to the market-share-weighted average unit price of the discounter’s PLs in the category (Ailawadi et al., 2003). Finally, to measure a NB manufacturer’s innovativeness, we counted the number of “innovations” that were launched in the category by the manufacturer in the German market (for the Aldi observations) and the Spanish market (for the observations from Mercadona) using data from Product Launch Analytics. Product Launch Analytics, formerly known as Productscan, is a subscription-based database that tracks CPG introductions (see, e.g., Sorescu & Spanjol, 2008, for an in-depth discussion on this data source). The vast majority of the NB manufacturer-category combinations in our sample are characterized by either no innovation (62% of the sample) or one innovation (23%) in the relevant category per year. Only 15% of the sample had more than one innovation per year in a given category.

4.3.4. Manufacturer market power

We measure manufacturer market power as the national “category share” of all of a NB manufacturer’s brands in terms of volume, using the panel data (Gielens & Steenkamp, 2007). Given that we interact this variable with PL production, we grand-mean-center it to facilitate interpretation (Irwin & McClelland, 2001).

4.3.5. Control variables

Apart from the above-stated variables, we also control for the differences between the two discounters by including a discounter dummy. Moreover, we control for PL production by the NB manufacturer for the discounter beyond the focal category (1 = yes, 0 = no PL production is observed in another category in our sample at that discounter). We expect a NB manufacturer to be more likely to produce PLs in the focal category if it also does so in one of the other categories. Moreover, we allow for PL production by the NB manufacturer in one category to also affect NB shelf presence in another category. On the one hand, one could expect that producing a PL in one category would also facilitate shelf presence in another category. On the other hand, as retailers embrace category management practices (e.g., Basuroy, Mantrala, & Walters, 2001), the pledge of PL production may not lead to goodwill in another category (for which a different category manager is responsible) or only to a lesser extent. Perhaps, if PL production in a category is already rewarded in that same category, it could even lead to a lower likelihood of shelf presence in another category.

In addition, we also control for the relative strength of the specific discounter in a given category. This factor may affect a discounter’s willingness to accept stronger, or to prefer weaker, NB opponents in its assortment. Moreover, it may affect a manufacturer’s willingness to produce PLs for the discounter in that category. To control for the relative strength of the discounter, we include the discounter’s (volume) share in the total (national) sales in the category. Given that we want to control for the discounter’s relative strength in a given category (compared to the other categories), we center this covariate within the discounter.
Finally, we include two product-class dummies (one to indicate household care and personal care, and one to indicate beverages, with food products as the base-level category) to control for unobserved category effects (see, e.g., Lamè et al., 2012 or Steenkamp, van Heerde, & Geyskens, 2010 for a similar practice). The covariates are not the focus of our study, but controlling for them provides a stronger test of our hypotheses (Greene, 2000).

We conducted formal tests to assess the skewness of all of the continuous covariates in our model. Following Tabachnick and Fidell (1996), we determined whether the skewness statistic exceeded twice the standard error of the statistic. Based on these tests, all of the continuous variables were found to be skewed and were log-transformed. None of the VIF statistics exceeded 3, suggesting that multicollinearity is not an issue (Cohen, Cohen, West, & Aiken, 2003).

### 5. Model

PL production is not a purely exogenous variable, but rather the result of strategic considerations from both the NB manufacturer’s and the discounter’s side. A failure to statistically correct for this endogeneity may not only lead to biased estimates but also to faulty conclusions about our key propositions (Hamilton & Nickerson, 2003). If manufacturers that produce PLs self-select themselves and/or are selected by the discounter on the basis of unobservable characteristics that affect both shelf presence and PL production, a problem of selection bias may arise. Selection bias can be corrected through the traditional two-step estimation technique that was proposed by Heckman (1979) or, alternatively, by using a Maximum Likelihood (ML) estimation on a system of equations. Because ML estimates are more efficient and have smaller standard errors than the two-stage estimates (Breen, 1996), we opt for that procedure.

To allow for the potential inter-correlation among observations (within a discounter) from multiple manufacturers within the same category and among observations from the same manufacturer across multiple categories, we use a robust two-way clustered-error term estimation on the system of equations (Cameron, Gelbach, & Miller, 2011).

Preliminary pooling tests failed to reject the assumption of homogeneity across both discounters ($\rho > .05$) for all but four effects: the impact of (i) manufacturer market power and (ii) its interaction effect with PL production on shelf presence, and the impact of (iii) price premium and (iv) PL production in a non-focal category on the PL-production decision. Hence, we pool the data across the two discounters, but we allow for discounter-specific effects in these four instances (along with a discounter-specific fixed-effects correction).

We estimate the following system of equations, where the first and the second equation are referred to as the outcome equation and the strategy-selection equation, respectively:

$$\text{PRESENCE}_{ijd} = \beta_0 + \beta_1 \text{D}_{\text{MERC}C} + \beta_2 \text{PLPROD}_{ijd} + \beta_2 \left( M_{\text{POWER}}_{ijd} \times D_{\text{ALDL}}_{ijd} \right) + \delta_2 \left( M_{\text{POWER}}_{ijd} \times D_{\text{MERC}}_{C} \right) + \epsilon_{ijd} \quad (1)$$

$$\text{PLPROD}_{ijd} = \delta_0 + \delta_1 \text{D}_{\text{MERC}C} + \delta_1 \text{GROWTH}_{ijd} + \delta_2 \text{EASE}_{ijd} + \delta_3 \text{ADV}_{ijd} + \delta_4 \left( \text{PREM}_{ijd} \times D_{\text{ALDL}}_{ijd} \right) + \delta_5 \left( \text{PREM}_{ijd} \times D_{\text{MERC}}_{C} \right) + \delta_6 \left( \text{INNOV}_{ijd} \times \delta_0 \text{M}_{\text{POWER}}_{ijd} \times D_{\text{ALDL}}_{ijd} \right) + \epsilon_{ijd} \quad (2)$$

with $D_{\text{MERC}}_{C}$ as a discounter-specific fixed effect and $Z_{ijd}$ as a vector that contains all of the other control variables (PL production in a non-focal category, discounter category share, and two product-class dummies), and $\Gamma$ and $\Omega$ the corresponding vector of coefficients in the outcome and strategy-selection equation, respectively.

In the outcome equation, $\text{PRESENCE}_{ijd}$ captures whether NB manufacturer $i$ obtained shelf presence in category $j$ with discounter $d$. $\text{PLPROD}_{ijd}$ reflects whether NB manufacturer $i$ produces PLs in category...
tics drive both shelf presence and PL production, no restrictions are
needed. To allow for the possibility that unobserved characteris-
tics drive the outcome equation (see, e.g., Guo & Fraser, 2010;
Lemke & Reed, 2001; Guo & Fraser, 2010) have already done so in another category, its acceptance probability in the
category remains increased \((b_1 + \gamma_1 = 1.53, p < .01)\), albeit
to a lesser extent. Hence, while production in the focal category still
creates additional goodwill, the incremental effect if the manufactur-
er already did so in another category is smaller (as evidenced by the
negative \(\gamma_1\)), as there is less signaling value to this additional pledge.
Similarly, when the manufacturer already produces in another cate-
gory, a refusal to do so in the focal category sends quite a negative sig-
nal to the discounter, and it jeopardizes the probability of gaining shelf
access from that discounter.

As for H2 and H3, no significant effect of manufacturer market
power is found for Aldi, neither in the produce condition \((b_2 + b_3 = .04, p > .10)\) nor in the no-produce condition \((b_2 = .07, p > .10)\).
For Mercadona, in contrast, we do find a significant effect of manufac-
turer power in the no-produce condition \((b_2 = .10, p < .01)\). In the
case of PL production, however, NB manufacturer power no longer has
a significant effect on the likelihood of obtaining shelf presence with
Mercadona \((b_2 + b_3 = .10, p > .10)\). Hence, manufacturer market
power diminishes the effect of PL production on the likelihood of NB
shelf presence, supporting H3. As for the control variables, we also
find a significant positive discounter-specific fixed effect for Mercadona
\((b_3 = 1.21, p < .01)\). This result is in line with the fact that Mercadona is
a soft discounter, and thus it is more inclined to accept NBs than Aldi,
which is a hard discounter. Finally, we find no effect of discounter cate-
gory share or of the product-class dummies on shelf presence.

6.1. Outcome equation estimates

Turning first to the outcome equation, we find that PL production
by a NB manufacturer for a discounter leads, at the average level of
manufacturer market power, to a significantly higher likelihood of
obtaining shelf presence in the category at that discounter \((b_1 = 2.26, p < .01)\). Hence, H1 is supported. Also when a manufacturer already
produces PLs in another category, its acceptance probability in the
category remains increased \((b_1 + \gamma_1 = 1.53, p < .01)\), albeit
to a lesser extent. Hence, while production in the focal category still
creates additional goodwill, the incremental effect if the manufactur-
er already did so in another category is smaller (as evidenced by the
negative \(\gamma_1\)), as there is less signaling value to this additional pledge.
Similarly, when the manufacturer already produces in another cate-
gory, a refusal to do so in the focal category sends quite a negative sig-
nal to the discounter, and it jeopardizes the probability of gaining shelf
access from that discounter.

As for H2 and H3, no significant effect of manufacturer market
power is found for Aldi, neither in the produce condition \((b_2 + b_3 = .04, p > .10)\) nor in the no-produce condition \((b_2 = .07, p > .10)\).
For Mercadona, in contrast, we do find a significant effect of manufac-
turer power in the no-produce condition \((b_2 = .10, p < .01)\). In the
case of PL production, however, NB manufacturer power no longer has
a significant effect on the likelihood of obtaining shelf presence with
Mercadona \((b_2 + b_3 = .10, p > .10)\). Hence, manufacturer market
power diminishes the effect of PL production on the likelihood of NB
shelf presence, supporting H3. As for the control variables, we also
find a significant positive discounter-specific fixed effect for Mercadona
\((b_3 = 1.21, p < .01)\). This result is in line with the fact that Mercadona is
a soft discounter, and thus it is more inclined to accept NBs than Aldi,
which is a hard discounter. Finally, we find no effect of discounter cate-
gory share or of the product-class dummies on shelf presence.

6.2. Strategy-selection equation estimates

We find evidence that a NB manufacturer’s sales growth negatively
affects the PL production decision \((b_2 = -.37, p < .05)\). Hence, we can
support H4. Ease of producing high-quality products in a given category
also has an impact. The easier it is to produce high-quality products, the
higher the likelihood of observing PL production by a NB manufacturer
for the discounter \((b_2 = 2.69, p < .10)\). As such, H5a is supported. This
finding provides support for the manufacturer’s rationale, as opposed to
the discounter’s rationale.

Concerning the marketing tools that are used by the NB manufactur-
er, we find that advertising expenditures have the expected negative ef-
fect. The likelihood that a NB manufacturer engages in PL production
decreases as it spends more on advertising \((b_2 = -.02, p < .10)\). Thus,
H6 is supported. As for H7, a higher price premium over hard discounter
Aldi’s PLs reduces a manufacturer’s probability of producing PLs
\((b_2 = -.80, p < .01)\). For soft discounter Mercadona, the price premium
does not have an effect. Concerning H8, we find that the innovativeness
of a NB manufacturer on the likelihood of PL production has a negative
effect \((b_2 = -.18, p < .10)\). As such, we find that the manufacturer’s ra-
tionale for PL production prevails over the discounter’s rationale.

To obtain insights into the relative importance of the three mo-
tives of PL production by a NB manufacturer (economic, quality/posi-
tioning, and brand-equity), we conduct a “what-if” analysis (see
Inman, Winer, & Ferraro, 2009 for a similar practice) that examines
the impact on a NB manufacturer’s PL-production propensity when
increasing each of the variables by one standard deviation above its
mean, while the other variables are held fixed at their baseline
levels (i.e., the mean for continuous variables and zero for dummy
variables). Among the marketing-mix variables (the brand-equity

---

16 While there is a general agreement on the importance of controlling for sample se-
lection issues, some authors (see e.g., Lemke & Reed, 2001; Guo & Fraser, 2010) have
warned against over-interpreting the actual size of the correction parameter (in our
setting, the correlation \(\rho\)).

17 Because of the skewness of the data, we use the mean and the standard deviation of the log-transformed variables.
motive), NB manufacturers dealing with Aldi are primarily concerned when a price premium is at stake (a one standard deviation increase in price premium reduces their propensity to produce with 11.8%, i.e., from 24.2% to 12.4%). Advertising and innovation-related concerns have a roughly equivalent, yet smaller, effect (advertising: −4.4%; innovation: −3.3%). Marketing mix concerns are much less of an issue (−2.6% and +2.4%).

The effect of manufacturer market power provides some additional insights: the likelihood of observing that a NB manufacturer will engage in PL production in a certain category increases the likelihood of producing in another. The effect is significantly positive for both Aldi ($\omega_A = .83, p < .05$) and Mercadona ($\omega_M = 2.69, p < .01$), although it differs in magnitude. Finally, the covariate that is related to discounter category share shows that the likelihood that a NB manufacturer will engage in PL production in a certain category increases when the discounter has a higher share of the market in that category ($\omega_D = .99, p < .01$), as there is more at stake for the manufacturer in its efforts to obtain shelf presence.19

### 6.3. Robustness checks

Several tests were performed to substantiate the robustness of our findings.

#### 6.3.1. Brand versus manufacturer power

Consistent with our conceptual framework, we measured a firm's market power by its total market share in a given category. However, a manufacturer's high market power can be derived through multiple moderately strong brands or through one very popular main brand. Retailers may be especially interested in listing the latter. We therefore re-estimated our model using brand rather than manufacturer power, with brand power defined as the volume share of the manufacturer’s largest brand.20 All of the results remain substantively the same.

#### 6.3.2. Absolute versus relative measures

In line with Ailawadi and Harlam (2004) and Slotegraaf and Pauwels (2008), we used absolute measures to operationalize the extent of a manufacturer’s advertising and innovation activity. However, some authors (see, e.g., Lamey et al., 2012; Reibstein & Wittink, 2005) have advocated the use of relative measures, given the importance of knowing how a manufacturer performs relative to competitors that operate under the same economic conditions. Therefore, we re-estimated our model, measuring the extent of a manufacturer’s advertising and innovations relative to the total level of all five leading NB manufacturers in that category. Our findings remain stable. The only exception is that the effect of the extent of advertising on category increases the likelihood of producing in another. The effect is significantly positive for both Aldi ($\omega_B = .83, p < .05$) and Mercadona ($\omega_B = 2.69, p < .01$), although it differs in magnitude. Finally, the covariate that is related to discounter category share shows that the likelihood that a NB manufacturer will engage in PL production in a certain category increases when the discounter has a higher share of the market in that category ($\omega_D = .99, p < .01$), as there is more at stake for the manufacturer in its efforts to obtain shelf presence.19

![Table 5](image_url)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>(t-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($b_0$)</td>
<td>−1.28</td>
<td>††† (−6.63)</td>
</tr>
<tr>
<td>PL production ($b_1$)</td>
<td>2.26</td>
<td>*** (11.25)</td>
</tr>
<tr>
<td>Manufacturer market power Aldi ($b_2$)</td>
<td>−.07</td>
<td>(−.46)</td>
</tr>
<tr>
<td>PL production * Manufacturer market power Aldi ($b_3$)</td>
<td>.11</td>
<td>(.56)</td>
</tr>
<tr>
<td>PL production * Manufacturer market power Mercadona ($b_4$)</td>
<td>−.61</td>
<td>(−4.97)</td>
</tr>
<tr>
<td>Mercadona ($b_5$)</td>
<td>.51</td>
<td>(5.86)</td>
</tr>
</tbody>
</table>

Control variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>(t-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discounter dummy (0 = Aldi, 1 = Mercadona) ($b_6$)</td>
<td>1.21</td>
<td>††† (5.93)</td>
</tr>
<tr>
<td>PL production in a non-focal category ($b_7$)</td>
<td>−.73</td>
<td>†† (−2.48)</td>
</tr>
<tr>
<td>Beverages dummy (food = baseline) ($b_8$)</td>
<td>.05</td>
<td>(.23)</td>
</tr>
<tr>
<td>Non-food dummy (food = baseline) ($b_9$)</td>
<td>−.02</td>
<td>(−.13)</td>
</tr>
</tbody>
</table>

Strategic selection equation (PL production)

| Intercept ($b_{10}$) | −3.35 | † (−1.74) |
| Manufacturer sales growth ($b_{11}$) | −.37 | ** (−1.91) |
| Ease of producing high-quality products ($b_{12}$) | 2.69 | † (1.70) |
| Manufacturer marketing tools | 0 | |
| Advertising ($b_{13}$) | −.02 | * (−1.43) |
| Price premium over PLs Aldi ($b_{14}$) | −.80 | *** (−3.45) |
| Price premium over PLs Mercadona ($b_{15}$) | .19 | (1.64) |
| Innovations ($b_{16}$) | −.18 | † (−1.67) |
| Manufacturer market power ($b_{17}$) | .21 | † (1.82) |

Control variables

| Discounter dummy (0 = Aldi, 1 = Mercadona) ($b_{18}$) | −1.16 | ††† (−5.07) |
| PL production in a non-focal category Aldi ($b_{19}$) | .83 | †† (2.22) |
| PL production in a non-focal category Mercadona ($b_{20}$) | 2.69 | †† (6.30) |
| Discounter category share ($b_{21}$) | .99 | †† (2.78) |
| Beverages dummy (food = baseline) ($b_{22}$) | −.37 | (−1.32) |
| Non-food dummy (food = baseline) ($b_{23}$) | −.27 | (−1.09) |
| Selection parameter ($\rho$) | −.87 | ††† (−2.62) |

Log-likelihood | −377.63 |

<table>
<thead>
<tr>
<th>Manufacturer’s rationale</th>
<th>Discounter’s rationale</th>
<th>Empirical findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer sales growth ($H_4$)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Ease of producing high-quality products ($H_5$)</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Manufacturer marketing tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advertising ($H_6$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price premium over PLs ($H_7$)</td>
<td></td>
<td>ns</td>
</tr>
<tr>
<td>Innovations ($H_8$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturer market power ($H_9$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For the manufacturer, a + (−) signifies the manufacturer is more (less) willing to produce PLs for the discounter. For the discounter, a + (−) signifies the discounter is more (less) interested to work with a particular leading NB manufacturer to produce its PLs. ‘ns’ means that no significant effect was observed.

---

18 We also examined whether the non-directional effects of ease of producing high-quality products and innovations are dependent on the level of manufacturer market power. Specifically, we added the interaction terms to Equation 2 (again taking into account whether these interactions could be pooled based on preliminary pooling tests). The interactions were not significant, neither individually nor jointly ($p > .10$).

19 We also tested for an interaction effect of PL production and discounter category share in the outcome equation. However, this effect was not found to be significant ($p = .33$). Moreover, all of the other results remained stable. For ease of interpretation and because we had no theoretical foundation for the interaction with discounter category share, we report the model with just the interaction of PL production and manufacturer market power.

20 Even though the correlation between brand and manufacturer power equals .89, we find some noticeable differences in power for some NB manufacturers. For example, while Campina’s largest brand has a market share of 3.5% in the German yogurt category, its total share in the category amounts to 7.3%. Its competitor, Ehrmann, in contrast, has a similar market share (7.9%), but with only one brand.
PL production, while similar in sign and magnitude, fails to reach significance (p = .19).

6.3.3. Timing of PL production and shelf presence

In our theorizing and empirical analysis, we implicitly assumed that PL production influences shelf presence (rather than the other way around). Hence, PL production should precede (or coincide with) NB shelf presence. Given the aforementioned secrecy surrounding PL production, it was impossible to identify the exact starting date of that activity. However, the temporal ordering of both events is not an issue for 398 (out of 450) manufacturer-category combinations (i.e., observations), as these NB manufacturers did not engage in PL production and/or did not obtain shelf presence in the category. Of the remaining 52 (12%) observations, where both PL production and shelf presence occurred, 30 (22) observations involved Aldi (Mercadona). For the 30 Aldi observations, we adopted an extensive procedure to eliminate any ambiguity surrounding the temporal ordering between both events as much as possible. For the 22 Mercadona observations, however, given prior agreements with our data provider that we would not contact the manufacturers, this was not possible. By comparing the date of the first PL-production documentation for Aldi (based on the above-listed sources, among which are Bertram, 2006; Schaab & Eschenbek, 2008; and Schneider, 2005, 2006) with the date that the NB manufacturers managed to obtain new shelf presence for at least one of their brands in the category (obtained from the GfK data), we observed that for ten of the observations, PL production unambiguously preceded the listing of a brand within the category. For the remaining 20 observations, we contacted the manufacturers. For seven observations, we could deduce from the manufacturers’ information that PL production definitely preceded shelf presence within the category, while in two instances, the reverse order was confirmed. For the remaining 11 observations, the manufacturers refused to reveal the required information.

To test the robustness of our insights to this timing issue, we dropped the two observations for which we obtained the confirmation of a reverse order. This action did not affect any of our substantive insights, with the only exception being that the effect of innovations became less significant (p-value of .17 rather than .09). In addition, to account for the fact that we could not unambiguously resolve the timing issue for 11 Aldi observations, we also implemented a more conservative test. Assuming that the same proportion (2/19) would apply to those 11 observations, we dropped an additional observation.21 We did this 11 times, each time dropping a different observation (on top of the two that were consistently excluded). The results are once more very robust. In every instance, the parameter for the PL-production dummy remains highly significant (p < .01) in the outcome equation, while the other findings also remain very stable. Looking at the median p-value across the 11 estimations, the same substantive conclusions as before are obtained, except for the effect of innovations, which was significant at a p-value of .09 in the main estimation but now obtains a p-value of .19. Across all of the 26 parameters that are listed in Table 5, the correlation between the original p-values and the median p-value that was obtained in this exercise was a high .98. Hence, even though the aforementioned secrecy surrounding PL production precluded us from obtaining the exact timing, we feel confident that this limitation did not substantively affect our results.

6.3.4. Shelf presence versus shelf breadth

We also explored whether our results remain stable when looking at the number of NBs with which a NB manufacturer obtains shelf presence, i.e., the extent of its presence, instead of only looking at whether a NB manufacturer was admitted. Manufacturers that offer better services to retailers may not only be rewarded with shelf presence per se (i.e., a 0/1 decision) but also with a more extensive presence (Mangold & Faulds, 1993). To that extent, we considered “shelf breadth,” which captures the number of NBs that are listed from a given manufacturer in a given category as an alternative dependent variable in the outcome equation. Again, using ML estimation, we estimated a recursive model with the treatment variable (PL production) as an endogenous (binary) regressor in the outcome equation, but the dependent variable in this outcome equation now takes the form of a continuous performance metric (see, e.g., Guo & Fraser, 2010; Jones, 2007). The major advantage of shelf breadth over shelf presence is that no discretization (which may involve an information loss) is needed. However, shelf presence has the attractive feature of being more stable over time, a non-negligible advantage given the cross-sectional (snapshot) nature of our study. In 47 out of the 185 manufacturer-category combinations, a NB manufacturer obtains shelf presence for more than one NB in a category, with a maximum of six brands. The results for the shelf breadth equation largely corroborate the shelf presence findings, with the exception of the effects of sales growth, ease of producing high-quality products, and manufacturer market power: while they are of the same sign, these variables now fail to reach statistical significance (p = .19, p = .30, and p = .36, respectively). Most importantly, however, we find that PL production, which is evaluated at the average level of manufacturer market power, positively affects shelf breadth in the category at the discountier (β1 = 1.53, p < .01). Hence, although it is not likely that multiple NBs from the same manufacturer will be accepted, the odds are better for a PL producing manufacturer than for a manufacturer that does not do so.

6.3.5. Including direct pull effects

We also tested for a direct ‘pull’ effect of the three marketing tools on shelf presence. No such evidence is found for a NB manufacturer’s price premium (p = .99) and innovations (p = .27). Only for advertising, a direct effect on the likelihood to obtain NB shelf presence is obtained (β1 = .02, p < .10). When allowing for potential pull effects, all of the other results are again very stable. The one change is that the effect of innovations in the production equation, while having the same sign as before, is no longer significant (p = .39).

6.3.6. Correlated errors across countries

We allowed for dependencies between observations using two-way clustering (i.e., for manufacturers and for categories within a country) in our main model. It is also possible that certain NB manufacturers are active in both countries. Ten such international NB manufacturers were found in our dataset (including Unilever and Nestlé). To allow for a potential correlation among observations from the same manufacturer across the two countries, we adopted an alternative two-way clustering approach (i.e., for categories within a country as before, but for manufacturers across both countries). Our results remain perfectly stable.

To summarize, across the various robustness checks, our results are found to be very stable, as detailed in Table 7. Foremost, the focal effect that PL production leads to a higher likelihood of obtaining shelf presence at discounters is stable across all of the robustness tests. Additionally, all of the other parameters in the outcome (shelf presence) equation consistently provide the same substantive insights. Not only the main effect of manufacturer market power, but also its interaction effect with PL production, and all of the control variables have similar effects across all of the checks. For the strategy-selection (PL production) equation, we achieve very robust results (with only one exception) for the effect of a manufacturer’s sales growth, the ease of producing high-quality products in a category, and a manufacturer’s market power. For the manufacturer marketing tools, we obtain equally robust results for a manufacturer’s price premium and for advertising. For the number of innovations, some variation in the significance of the effect is found, even though the

21 For 19 (=10 + 7 + 2) observations, the ambiguity could be resolved.
The no shelf presence was obtained for any of their brands. Given that
ket for close to 60% of all of the manufacturer-category combinations,
cused on the leading manufacturers in the German and Spanish mar-
channel is far from straightforward or automatic. Although we fo-
(IGD, 2005, p. 34), they should realize that gaining access to that
discount sector as
and discounters. First, while many NB
manufacturer should engage in PL production is, however, not an easy
ight of the current economic downturn. The decision of whether a NB
come increasingly attractive to CPG manufacturers, particularly in
light of the potential downsides to such a dual-branding strategy. Foremost,
concerned with the potential downsides to such a dual-branding
strategy. (de Jong, 2007). In this article, we empirically investigated the determinants, as well as the effect on discounter good-

cs, such as the payment of slotting allowances (Sudhir & Rao, 2006)
and/or increased trade promotions (Narasimhan, 2009), are rarely
used by discounters (PlanetRetail, 2010).

7. Discussion

Faced with soaring PL shares and volumes, PL production has be-
come increasingly attractive to CPG manufacturers, particularly in
light of the current economic downturn. The decision of whether a NB
manufacturer should engage in PL production is, however, not an easy
one. Hence, it is not surprising that most NB manufacturers are still
struggling with the issue (de Jong, 2007). In this article, we empirically
investigated the determinants, as well as the effect on discounter good-
will, of NB manufacturers’ engagement in PL production. We tested our
hypotheses on a sample of 450 NB manufacturer-category combina-
tions for two leading discounters in the German and Spanish market,
using a uniquely assembled dataset that combines extensive field re-
search with various secondary data sources.

Our analyses offer important insights to both NB manufacturers and
discounters. First, while many NB manufacturers now regard the
discount sector as “an essential channel they proactively approach”
(IGD, 2005, p. 34), they should realize that gaining access to that
channel is far from straightforward or automatic. Although we fo-
cused on the leading manufacturers in the German and Spanish mar-
ket for close to 60% of all of the manufacturer-category combinations,
no shelf presence was obtained for any of their brands. Given that

managers of convenience goods strive to distribute their brands as in-
tensively as possible (Coughlan, Anderson, Stern, & El-Ansary, 2001,
p. 288), missing out on some of the fastest-growing retailers repre-
tsents a substantial set-back. To make matters worse, discounters
(especially hard discounters) tend to be influenced less by the manu-
facturers’ power status than conventional supermarkets; additionally,
the tools that manufacturers are most familiar with to ensure shelf ac-
cess, such as the payment of slotting allowances (Sudhir & Rao, 2006)
and/or increased trade promotions (Narasimhan, 2009), are rarely
used by discounters (PlanetRetail, 2010).

However, we find strong support for the (thus-far untested) notion
that PL production for a discounter in a given category substantially in-
creases the likelihood of obtaining shelf presence in that category.
While quite some NB manufacturers are still reluctant to produce PLs,
they should be aware that a refusal to engage in PL production for a
discounter may seriously jeopardize their chances of gaining access to
the limited spots on that discounter’s shelves. Moreover, unlike the
aforementioned tools (such as slotting allowances and promotional
support), which entail a direct monetary transfer from a manufacturer
to a retailer (Scott Morton & Zettelmeyer, 2000), PL production, when
instrumental in securing NB shelf access, contributes positively to a
manufacturer’s bottom line in various ways: (i) revenues from the
NB(s) that are sold through the discount channel, (ii) revenues from the
PLs (which could potentially take place through otherwise unused,
and therefore costly, over-capacity), and (iii) positive margin implica-
tions that may result from the more intense relationship with the retail-
er (ter Braak et al., 2013).

In spite of these potential benefits, many managers remain
concerned with the potential downsides to such a dual-branding
(PL production along with NB development) strategy. Foremost,
they fear the reaction of their customer base if it were to become

 Detailed results are available from the first author upon request.
known that they also produce (cheaper) PLs. This fear is reflected not only in the lower PL-production likelihood that we obtained for manufacturers that currently charge a higher premium for their brands (especially when dealing with very price-oriented hard discounters), and/or that invest heavily in advertising in an effort to differentiate their brands, but also in the efforts that those that do produce take to conceal this practice. Many companies go even one step further, and separate both activities not only commercially, but also in the location of their physical production facilities. Firms as Reckitt Benckiser (a leading manufacturer of household products) and McCain (manufacturer of, among other categories, frozen potato products), for example, have completely separated all of their PL activities from those for their own brands, to the extent of creating separate companies (de Jong, 2007). Not only does this restrain the flexibility on how to utilize the total production capacity (Bergès & Bouamra-Mechemache, 2011), manufacturers should also be careful that this strategy does not hurt potential relational (goodwill-related) benefits.

Apart from these customer-equity concerns, we find that manufacturers are also afraid that they will become pressured to share their latest technologies and innovations with the retailer/discounter, which becomes even more likely when only a few players can assure high-quality production. In spite of this concern, PL production can be an attractive opportunity for innovation-oriented manufacturers as well. Apart from the aforementioned financial benefits, they can use such production as “a chance to try out product ideas at much lower costs” (Dunne & Narasimhan, 1999, p. 42; ter Braak et al., 2013). Indeed, launching a new product is risky and costly. When manufacturers are not sure if a new product will work (which is the case for many new products), they can limit their risk through an upfront involvement of a discounter (while also bringing shelf presence for their NBs into the negotiations) and subsequently supply a product that primarily substitutes for their competitors’ NBs (Dunne, 1999).

Finally, many manufacturers operate in multiple categories and may be more/less reluctant to engage in PL production in some categories than in others, if only because their advertising intensity differs across categories or because they have more intellectual property to protect in some categories than in others. Such firms should be aware of the fact that the impact of PL production in one category (which they themselves may perceive as an initial pledge towards a better relationship with the discounter) may actually amplify the negative implications of refusing to do so in other categories.

Our analyses also offer some key insights for discounters. Given their historically almost exclusive focus on PLs, they may have very limited knowledge on (or even interest in) NB manufacturers. Deleersnyder et al. (2007) have discussed that it may pay off for discounters to monitor the NB scene more closely to identify which brands to add to their assortment to maximize their overall category performance. However, apart from these more demand-driven considerations, it may also be useful to identify which NB manufacturers will be most eager to obtain shelf space for their NBs, and/or least reluctant to engage in PL production, as this will reflect on their bargaining position when negotiating supply terms. Declining sales for the other party’s NBs, for example, may signal over-capacity, and a higher willingness to not only produce PLs, but also to do so at lower wholesale costs. Meanwhile, when looking to add more premium PLs to the assortment (which involves vertical and horizontal differentiation from existing alternatives; Kumar & Steenkamp, 2007), they may primarily want to approach manufacturers that are known for their innovative capabilities. To overcome these manufacturers’ initial reluctance to do so, offering shelf presence for their NBs may be an appealing option.

7.1. Limitations and further research

In line with Kaufman et al. (2006) and Ailawadi, Pauwels, and Steenkamp (2008), we considered two different retailers to test our hypotheses. It was encouraging to see that most of our findings generalized across both discounters, even though they have a different positioning (hard vs. soft discounter), and they come from two different countries (Germany vs. Spain). Moreover, the observed differences (i.e., the role of manufacturer market power in the outcome equation and the role of price premium in the selection equation) had considerable face validity in light of the discounters’ different positioning. Still, it would be interesting to investigate to what extent our findings can be generalized to (i) other product categories, (ii) other countries, and (iii) other retail formats. Additionally, the recent economic crisis may contribute both to further growth (and hence, a stronger negotiation position) for the discounters, while NB manufacturers may become more inclined to engage in PL production.

Because discounters operate with a limited assortment, shelf space that is given to NBs is presumably the most important dimension of retailer goodwill. For traditional retailers, shelf space might not be such a good measure of goodwill. To them, slotting allowances provided by NB manufacturers play a key role in a retailer’s decision to carry a NB (Sudhir & Rao, 2006), while they are rarely used in a discounter setting (PlanetRetail, 2010). Future research could investigate what extent our findings are generalizable to traditional retail formats on other dimensions of retailer goodwill, such as the extent of pass-through of trade promotions (Ailawadi & Harlam, 2009), the size of the slotting allowances (Sudhir & Rao, 2006), a NB manufacturer’s influence on a retailer’s promotional calendar (Dunne & Narasimhan, 1999), and/or the likelihood of being appointed “category captain” (Subramanian et al., 2010).

Similar to previous studies (see, e.g., Nijs, Dekimpe, Steenkamp, & Hanssens, 2001; Pauwels, Srinivasan, & Franses, 2007), we focused on the top players in each category to keep the data-collection effort manageable. Due to our focus on the top-five NB manufacturers in the category, the more limited variation in the manufacturer market-power variable may have contributed to the non-significant effect of that variable for Aldi. Future research could also investigate non-top-five manufacturers. Our data are further limited in that it offers only a historical snapshot. While we took great care to ensure that PL production precedes shelf presence, we did not have precise information on the starting dates of both events. A longitudinal dimension would undoubtedly offer additional insights.

Furthermore, future research could investigate potential cross-country effects. For example, a top NB manufacturer could supply PLs to a discounter in a country in which the NB manufacturer is not active. Such PL production would create goodwill with the focal discounter, while the negative implications (reputational and competitive disadvantages) for the NB manufacturer would be limited. In addition to these cross-country effects, future research could also investigate cross-retailer effects of NB manufacturers’ PL production decisions. IGD (2005) reported that a concern that was mentioned by PL suppliers for the discount channel was how this engagement would impact their relationship with other retailers. Although the goodwill of the discounter for which the NB manufacturer starts to produce may increase, this could come at the expense of deteriorated relations with other retailers. Similarly, should PL production be restricted to a single discounter? Such exclusivity could create even more goodwill with the focal discounter, but the negative implications with other retailers (discounters) may then become more pronounced. Finally, while we provided initial evidence that the implications of PL production may extend beyond the focal category (through the γ parameter), it may be interesting to quantify the net implications across all of the categories in which a manufacturer is active. In doing so, it would pay off to also consider the moderating influence of a manufacturer’s power across all of the categories.

23 In addition, the economic situation may have been different, given the later year of the Spanish data collection. We thank an anonymous reviewer for this observation.
We have only begun to scratch the surface of research possibilities in an area that warrants more attention. We hope that this article will provide a stimulus for more research on PL production.

Acknowledgments

We are grateful for the constructive comments that we received from the anonymous JIRM reviewers. We would also like to thank Karen Gedenk, Els Gijsbrechts, Koen Pauwels, Raj Sethuraman, and Jan-Benedict Steenkamp for their insightful comments on an earlier draft. We further acknowledge AMark for providing access to the GfK consumer panel data and thank Kantar Worldpanel for their help in obtaining the private label supplier information for Mercadona. Finally, we are indebted to the Netherlands Organization for Scientific Research (NWO) for its financial assistance.

References


So you want to delight your customers: The perils of ignoring heterogeneity in customer evaluations of discretionary preferential treatments

Raphaëlle Butori *, Arnaud De Bruyn
ESSEC Business School, Avenue Bernard Hirsch, 95000 Cergy, France

**A R T I C L E  I N F O**

Article history:
First received in 9, June 2011 and was under review for 9 months
Available online 20 June 2013

Guest Editor: Marnik G. Dekimpe
Area editor: Werner J. Reinartz

Keywords:
Discretionary preferential treatment
Preference heterogeneity
Justification
Imposition
Visibility
Surprise
Need for distinction
Negotiation proneness

**A B S T R A C T**

Many firms assume that customers like to feel special and to receive discretionary preferential treatments (DPT). This research argues that the reality is more complicated: the same preferential treatment may delight one customer but enrage or embarrass another. To help companies align their DPT with their customers’ preferences, this article identifies four dimensions along which consumers positively or negatively evaluate DPT: justification, imposition, visibility, and surprise. This article then introduces customer heterogeneity in the form of two individual traits that moderate DPT evaluations. Through two studies, the article shows that distinction seekers prefer visible rewards that impose on other customers, but negotiators prefer unjustified, non-surprising privileges. Finally, by tying consumer preferences to two readily available variables (age and gender), this article concludes with a set of practical guidelines for the companies that hope to align their DPT strategy with customer profiles.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

One of the authors went to a fancy restaurant with a friend to celebrate a special event. Because the restaurant owner personally knew one of them, the staff went to great lengths to please them: they were seated at a central table, received exquisite attention and lavish service, and were offered special dishes that were not on the menu. The benefits were so great that the two became the center of attention (and envy) of the entire restaurant. The author was so embarrassed that she swore never to go to that restaurant again; her friend, instead, was delighted.

On another occasion, one of the authors bought an expensive piece of luggage in an airport store just before entering a long security queue in which hundreds of passengers were waiting. A store employee offered him a note to hand to the airport security personnel, which allowed him and his wife to proceed through the handicapped aisle and skip a 90-minute wait in line. His wife was thrilled by this special treatment. When he recalls walking down the handicapped aisle, though, bypassing hundreds of passengers, he refers to it as “the walk of shame.”

Both examples are typical illustrations of preferential treatments that backfire and trigger feelings of guilt and embarrassment. These examples raise a key question about the efficient use of preferential treatments: what type of preferential treatments should privileged consumers receive? We examine this question in relation to a specific, mostly overlooked type of preferential treatment, namely, non-contractual preferential treatments. That is, existing research mostly considers preferential treatments in the context of a contractual reward process, involving loyalty programs with explicitly stated rules and policies (e.g., Drèze & Nunes, 2009; Kivetz & Simonson, 2003; Nunes & Drèze, 2006; Roehm, Pullins, & Roehm, 2002). However, some preferential treatments are granted at the whim of the company, which alone determines the recipients, nature, and value of the rewards (Kumar & Shah, 2004), often at the discretion of its frontline employees. For example, ACCOR hotels’ desk managers offer non-contractual privileges, such as room upgrades, free breakfast, or dedicated parking spaces to selected clients in addition to the corporate privileges offered by the ACCOR loyalty program. These non-contractual forms of preferential treatment are discretionary preferential treatments (DPT), which we define as the selective granting of non-contractual advantages to a limited number of customers. In essence, DPT (a) is selective, (b) comes in addition to contractual preferential treatment, (c) involves an informal granting process (i.e., does not rely on publicly stated rules and policies), and (d) allows for the decision flexibility of the front-line employees.

Unlike contractual preferential treatment, DPT offers various advantages that make it an interesting managerial tool. Because its rules are not publicly stated, DPT (1) cannot produce liabilities such as ongoing obligations to recipients (Shugan, 2005), (2) eliminates...
the potential for demotions to lower levels of service and their negative consumer outcomes (Wagner, Henning-Thuraus, & Rudolph, 2009), and (3) increases customization flexibility, which can stimulate long-term loyalty (Shugan, 2005). In addition, because DPT is not just a function of the volume purchased, it can be used to treat selected customers even better, thereby stimulating a feeling of being treated as special (O’Brien & Jones, 1995). Finally, because frontline employees have more latitude to grant it, DPT strengthens the employee–customer relationship, which stimulates customer share, price premiums, and sales growth (Palmatier, Scheer, Houston, Evans, & Gopalakrishna, 2007).

Despite these benefits and its managerial relevance, little research considers how consumers value DPT. This research gap is problematic because consumers’ reactions to DPT are heterogeneous. The same DPT, such as being favored by restaurant staff or allowed to cut a long waiting line, might delight one customer but embarrass another. If companies ignore the heterogeneity in customers’ preferences for DPT, they might offer rewards that are not valued by the targeted customers—or worse, that elicit negative reactions—and squander valuable marketing resources (Reinartz & Kumar, 2000). Firms thus must ask the question that guides our research: what type of DPT should be offered and does the answer vary predictably across consumers? To answer this question, we organize this manuscript as follows.

In the first section, we develop the theoretical underpinnings for this research. We identify four key dimensions along which customers evaluate DPT: justification (i.e., whether DPT is warranted by an existing relationship between the firm and the customer), imposition (whether DPT detrimentally affects other customers), visibility, and surprise. Building on equity theory (Adams, 1965), we argue that most people prefer DPTs that are justified, non-imposing, non-visible, and surprising. In addition, we rely on social comparison theory (Festinger, 1954) to hypothesize that these general preference tendencies are moderated by the consumer’s need for distinction and negotiation proneness; consumers who like to be distinguished from others prefer imposing and visible DPT, whereas consumers who prefer to negotiate favor unjustified, non-surprising DPT.

In the subsequent two sections, we report the results of two studies that were run in the context of a hotel restaurant (Study 1) and a retail store (Study 2), in which we formally test the hypotheses. Most of the main effects and moderators receive confirmation. We also show that the reported heterogeneity in customers’ preferences for various DPT can be partly anticipated by two readily available variables: age and gender. Building on these findings, we provide a set of practical guidelines for companies and conclude with some suggestions for further research.

2. Theoretical framework

2.1. Dimensions of DPT evaluations

Discretionary preferential treatments provide non-contractual advantages to a limited number of customers. Unlike their contractual counterparts (e.g., rewards earned through loyalty programs), DPTs entail (1) a greater degree of distinction between customers, in that they appear in addition to contractual rewards (Kumar & Shah, 2004), and (2) a discretionary nature, such that their granting process does not rely on preexisting rules or conditions. These differences suggest four dimensions of DPT that are particularly worthy of investigation.

First, because DPTs are added on to contractual preferential treatments, but companies’ resources are limited, they may mandate smaller resource allocations to non-privileged consumers to allow more resources to be devoted to the privileged ones (Kamakura, Mittal, de Rosa, & Mazzon, 2002). DPTs can thus be granted to the detriment of non-privileged consumers, which makes imposition on others the first dimension worthy of investigation.

Second, the DPT process is informal, such that DPT may be granted arbitrarily. This potential for arbitrary decisions raises the question of DPT justification, that is, whether the DPTs are warranted by the nature of the relationship between the customer and the company.

Third, the discretionary DPT process also allows the frontline employees to make DPT decisions on the fly and possibly in front of an audience of non-privileged customers. This social setting enables both the privileged and the non-privileged customers to compare what they receive with what others receive, thereby making visibility another central dimension of DPT.

Fourth and finally, because DPTs do not rely on publicly stated rules and conditions, they leave room for the unexpected and have the potential to create delightful experiences. Surprise represents the fourth dimension of DPT that we study.

In turn, we use these four dimensions to define the type of DPT that consumers encounter. With equity theory (Adams, 1963), we predict the customers’ general preferences (e.g., whether most customers prefer surprising or unsurprising DPT). We then build on social comparison theory (Festinger, 1954) to introduce moderators that mitigate these general tendencies.

2.2. Preferences for DPT dimensions

2.2.1. Equity theory

According to equity theory (Adams, 1965), participants in social exchange relationships compare their outcomes from the exchange with their inputs into it (internal equity) as well as the balance between their own outcome/input ratio and those of significant others (external equity). If the outcome/input ratios of partners appear unequal, inequity exists. The greater the inequity (over- or under-reward), the more distress the participant feels.

In a consumption setting, under-rewards tend to create feelings of resentment (Lapidus & Pinkerton, 1995), whereas over-rewards prompt the suspicion that companies are employing manipulation tactics to induce specific behaviors (e.g., encourage spending). The perception of such manipulative intent may result in a boomerang effect, whereby the consumers reject the encouraged behavior (e.g., Clee & Wicklund, 1980). Either way, people tend to prefer situations that they perceive to be equitable.

Because DPTs establish unequal levels of treatment among customers, they drive both privileged and non-privileged customers’ attention toward the perceived inequity of their rewards. Equity theory (Adams, 1965) is thus a particularly relevant framework for understanding how consumers evaluate the four dimensions of DPT.

2.2.2. Justification

In general, justification refers to the presence or absence of any valid grounds for an act or course of action. Because DPT generally falls within the scope of a relationship between the firm and the customer (Gwinner, Gremler, & Bitner, 1998), it is justified (unjustified) when it has been warranted (not warranted) by the nature of their relationship. Unjustified DPTs create an imbalance in the consumers’ outcome/input ratios, such that they should generate more distress than justified DPT (Adams, 1965).

Unjustified DPT might also suggest that the company is attempting to induce specific behaviors (e.g., buy more expensive items than planned). For instance, a consumer offered a free drink at his or her first visit to a restaurant might experience an undesirable feeling of indebtedness. Such preferential treatment might be appreciated, but it would have created a more positive feeling had this suspicion not been aroused. Therefore, unjustified DPT could not only generate distress but could also result in consumer inferences of manipulative intent by companies. In a consumption setting, awareness of this manipulative intent generates negative reactions, such as irritation.
(e.g., Edwards, Li, & Lee, 2002). Accordingly, we hypothesize the following:

**H1.** The more justified the DPT, the more positively the privileged consumers evaluate it.

2.2.3. Imposition

To issue DPT, the firm must allocate its limited resources to a limited number of its consumers (Bolton, Lemon, & Verhoef, 2004). This allocation implies that marketing efforts for non-privileged customers must be reduced (Kamakura et al., 2002), which can negatively influence the perceived level of service provided to those customers. For example, if a privileged customer monopolizes the attention of a sales clerk in a crowded store, the non-privileged customers lose their opportunity to ask questions and seek advice. From an equity theory perspective (Adams, 1963, 1965), imposing DPT therefore lowers the non-privileged consumers’ outcomes, which creates an imbalance in their outcome/input ratios. Consistent with the internal equity principle, this imbalance should result in feelings of unfairness among non-privileged consumers, as well as possible hostile reactions to privileged consumers (Gwinner et al., 1998).

Imposing DPT creates imbalances not only within the non-privileged customers’ outcomes and inputs but also between privileged and non-privileged customers’ outcome/input ratios. According to the external equity principle, this imbalance should generate perceptions of unfairness among both the privileged and the non-privileged consumers, who feel over-rewarded and under-rewarded, respectively. Because over-reward generates distress and negative feelings, such as guilt (Hassebrauck, 1986; Steenhaut & Van Kenhove, 2005), imposing DPTs may also put the privileged customers in an uncomfortable situation.

In summary, when receiving imposing DPTs, the consumers may feel embarrassed in the face of negative reactions from non-privileged customers, and they may also experience distress and guilt. Should they be given a choice between an imposing DPT and an unimposing DPT with the same benefits, we expect that these customers would prefer the unimposing DPT. All else being equal, the negative impact of DPT on non-privileged consumers should reduce its subjective value:

**H2.** The more imposing the DPT (i.e., the greater its negative impact on the level of service received by non-privileged customers), the more negatively the privileged consumers will evaluate it.

2.2.4. Visibility

Research on preferential treatment suggests a mixed impact on several relational variables. Lacey, Suh, and Morgan (2007) offer strong support for the use of preferential treatment as a relationship marketing tool, showing that it increases sales, customer share, word of mouth, and feedback. Yet Hennig-Thurau, Gwinner, and Gremler (2002) find no significant relationship between preferential treatment and either satisfaction or loyalty, and only a modest relationship with word of mouth. In a similar vein, De Wulf, Odekerken-Schröder, and Iacobucci (2001) find no significant effect of preferential treatment on perceived relationship investments. To account for this finding, these authors note that some people might feel embarrassed when they are openly favored in front of others, which raises the idea that the visibility of the privileges (i.e., granted in private versus in public) influences how consumers evaluate the DPT (Melnyk & Van Osselaer, 2012). Consistent with equity theory (Adams, 1965), we posit that the visibility of privileges reinforces the comparison processes, such that others notice the increase in the privileged customers’ outcomes. Because the underlying justifications for this increase may not be known (e.g., loyalty, amount of last purchase), the publicly privileged consumers may be perceived as being unfairly privileged, leading to embarrassment and guilt. We therefore expect the following result:

**H3.** The more visible the DPT, the more negatively the privileged consumers evaluate it.

2.2.5. Surprise

Unlike contractual preferential treatments, DPTs do not rely on explicitly and publicly stated rules and policies. Thus, consumers may not know in advance whether they will be privileged. Should they expect special treatment as a reward for repeated patronage (consistent with the internal equity principle; Adams, 1963, 1965), the content of this special treatment (i.e., advantages) would still remain unknown because no publicly stated rules determine the conditions and nature of the DPT rewards. This absence of contractual terms leaves room for the unexpected and thus the potential for delight (Rust & Oliver, 2000).

Contrary to “musts,” which are the central features of an offer, and “satisfiers,” which are embellishments to the basic offer, “delights” are attributes that the consumers do not expect to find in the offer (Kano, Seraku, Takahashi, & Tsuji, 1984; Oliver, 1997). Delights are surprising in nature, and their impact on satisfaction is always positive. In a growing culture of entitlement (Boyd & Helms, 2005), in which customers who know their worth expect special privileges that reflect it, the potential for delight constitutes a powerful differentiating tool. Because DPTs hold this potential, we expect surprise to be a dimension of DPT that consumers attend to carefully; thus, we propose the following:

**H4.** The more surprising the DPT, the more positively the privileged consumers evaluate it.

2.3. Moderators

Several studies suggest that consumers vary in their sensitivity to relationships marketing practices. For example, De Wulf et al. (2001) find that companies’ efforts to enhance relationships with regular customers are not always positively perceived. The impact of these efforts on relationship quality depends on the consumer’s product category involvement (Mittal, 1995) and relationship proneness. Butori (2010) also shows that consumers vary in their receptiveness to the symbolic, hedonic, and utilitarian benefits of DPT. Whereas some consumers are delighted at the seller’s special attention, others do not value the feeling of distinctiveness. These findings suggest that no matter what the firms do to please their customers, the effects will be tempered by the individual consumer’s characteristics (Bendapudi & Berry, 1997; Christy, Oliver, & Penn, 1996; Day, 2000).

A wide range of individual variables likely influence how consumers evaluate DPT. We focus on those variables related to the core, essential characteristics of the DPT, namely, its selective and discretionary process. Because the DPT process is selective, it establishes a distinction among customers; because it is discretionary and not based on a well-defined set of rules, it leaves room for negotiation. People vary in their sensitivity to distinction (Brewer, 1991) and are differentially likely to engage in negotiations (e.g., Harris & Mowen, 2001; Rubin & Brown, 1975). Thus, a consumer’s need for distinction, defined as the degree to which being distinct from others is important to the consumer’s self (White & Argo, 2011), and negotiation proneness, defined as a consumer’s desire to engage in negotiations (Mowen & Spears, 1999), likely influence how consumers evaluate the essential characteristics of DPT. We use these variables in our attempt to explain heterogeneity in DPT evaluations, and in the next sections, we build on the social comparison theory (Festinger, 1954) to explain how the need for distinction and negotiation proneness likely moderate DPT evaluations.

---

1 Other individual variables, such as locus of control, narcissism, assertiveness, or market mavenism, might also influence how consumers evaluate DPT. We leave these effects for further research.
2.3.1. Need for distinction

Social comparison theory (Festinger, 1954) is grounded in three propositions: (1) people evaluate their own abilities and opinions; (2) in the absence of any objective benchmarks, people turn to social comparisons (i.e., with others); and (3) whenever possible, people make comparisons with similar others.

Several types of social comparisons are possible, depending on whether the person undertakes a comparison with someone who is superior in some way (upward social comparison) or inferior or less fortunate (downward social comparison) than him- or herself. Whereas downward comparisons enhance subjective well-being (Affleck & Tennen, 1991; Gibbons, 1986; Wills, 1981, 1991), upward comparisons lead to negative affective reactions, especially if the focal comparison dimension is significant for self-esteem (Richins, 1991). In this case, comparisons even pose a threat to self-esteem and have an ego-deflating effect.

People are not equally prone to engage in these social comparisons, and the impact of comparisons on perceived self-worth varies from one individual to another. Because the need for distinction is the expression of more general narcissistic tendencies (American Psychiatric Association, 2000) and because narcissists are overly dependent on social sources for self-affirmation (i.e., they tend to orchestrate downward social comparisons to maintain positive self-views, Horvath & Morf, 2010), distinction seekers should be particularly prone to engage in social comparisons. However, these distinction seekers are not genuinely concerned with others; rather, they use social interactions as a stage for maintaining a positive view of themselves and to elicit the admiration of others (Besser & Zeigler-Hill, 2010). Accordingly, the distinction seekers should be particularly sensitive to comparisons made in public but absolutely not concerned about the impact of their privileges on others. These people may even be flattered by visible and imposing DPTs because such privileges signal their importance. Therefore, we hypothesize the following:

**H5.** The higher the consumer’s need for distinction, the more positively he or she evaluates an imposing DPT.

**H6.** The higher the consumer’s need for distinction, the more positively he or she evaluates a visible DPT.

2.3.2. Negotiation proneness

People engage in negotiations in response to two primary motivations: economic, such that they hope to receive a tangible value associated with the outcome of the negotiation; and non-economic, resulting from the pleasure associated with demonstrating negotiation competence (Assor & O’Quin, 1982; Rubin & Brown, 1975). Because DPTs do not rely on preexisting conditions that determine which consumers should be privileged and how, consumers might take the initiative and ask for a favor and negotiate its content. In this sense, DPTs provide a means for the consumers to demonstrate their competence. This empowerment increases the opportunity for social comparisons (Wathieu et al., 2002) and strengthens the impact of the comparisons on subsequent affective reactions (e.g., increased or decreased self-esteem). Indeed, downward comparisons enhance well-being primarily when the comparison is esteem-relevant or when people perceive their superior standing as stable and within their control (Major, Testa, & Bylsma, 1991). In other words, a downward comparison is more ego-inflating when the person has control over the skill being evaluated. With DPT, the consumer’s attribution of the preferential treatment to his or her negotiation skills therefore influences the magnitude of the positive effects on self-esteem. The more challenging the negotiation, the greater the level of competence demonstrated, and the greater the pride associated with the negotiation (Rose, 1988; Schindler, 1998). Because unjustified DPTs are more difficult to negotiate than justified DPTs (i.e., they do not rely on existing loyalty or extant relationships), they are more difficult and challenging to obtain, which makes them more valuable to those people prone to negotiate. Accordingly, we hypothesize the following:

**H7.** The higher the consumer’s negotiation proneness, the more negatively he or she evaluates a justified DPT.

People who are prone to negotiating should also be particularly willing to exert control over the DPT negotiation process and take the initiative to ask for preferential treatment. Because DPTs granted by surprise prevent these customers from playing an active part in the process, we propose the following:

**H8.** The higher the consumer’s negotiation proneness, the more negatively he or she evaluates a surprising DPT.

Table 1 summarizes our hypotheses, which we test in two different managerial contexts.

3. Study 1: hotel restaurant

3.1. Experimental setting

In this study, we asked respondents to imagine being the privileged customer of a hotel where they are staying. When they arrive at the hotel’s restaurant, a table has been set aside for them. We used a hotel scenario for two reasons: (1) hotels represent a context in which consumers regularly experience, and (2) DPTs in this business often vary across the dimensions of interest. The experiment presented four hypothetical DPTs at a time, in a two-by-two table, and the respondents ranked them by decreasing order of preference, from 1 (most preferred) to 4 (least preferred), such that the measure featured trade-offs similar to those commonly used in conjoint studies (Carroll & Green, 1995; Green & Srivinwasan, 1978). The respondents also indicated their confidence in their top choice on a seven-point Likert scale.

Each DPT was described by two of the four evaluation dimensions (justification, imposition, visibility, or surprise). The questionnaire contained all possible two-by-two combinations, for a total of six tables, each with four cells, and therefore 24 data points per respondent. Table 2 (left) provides the descriptions of each manipulated dimension.

To test the moderating influences of the need for distinction and negotiation proneness, we asked the respondents to complete Butori’s (2010) need for preferential treatment (NPT) scale. This scale measures a consumer’s receptivity to a DPT. In particular, its distinction subscale measures the consumer’s receptivity to the symbolic benefits associated with the distinction that a DPT establishes (e.g., “I do not like to feel like any customer”), and its play subscale measures the consumer’s receptivity to the fun benefits associated with a DPT negotiation process (e.g., “I enjoy negotiating advantages as much as actually using them”). These subscales therefore capture the extent to which consumers like to be distinguished from others and play the game of negotiation, respectively. The subscales serve to identify distinction seekers and negotiators (Appendix A provides the items as well as the factor loadings and Cronbach’s alphas obtained for the survey).

Table 1

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Need for distinction</th>
<th>Negotiation proneness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Justification</td>
<td>$H_5$</td>
<td>$H_7$</td>
</tr>
<tr>
<td>Imposition</td>
<td>$H_2$</td>
<td>$H_6$</td>
</tr>
<tr>
<td>Visibility</td>
<td>$H_4$</td>
<td>$H_8$</td>
</tr>
<tr>
<td>Surprise</td>
<td>$H_3$</td>
<td>$H_6$</td>
</tr>
</tbody>
</table>
3.2. Pretest

Before administering the questionnaire, we conducted a pretest with 50 students to ensure that each description correctly manipulated its assigned modality. The respondents rated each description level for each dimension (Table 2); for example, after reading the description of the visible modality, the participants rated the preferential treatment as visible or not visible on a seven-point semantic differential scale. This pretest was conclusive: each description correctly manipulated its assigned modality (details available on request).

3.3. Sample

The 125 respondents in the main sample completed the study by filling in the six tables and the NPT scale. These respondents were all participants of a five-day seminar on tax law. To reduce any carry-over effects, they filled in the tables on the first day of the seminar and the NPT scale on the fifth day (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The content of the seminar had nothing to do with the purpose of the study, so no contamination should intervene between the measures of the two sets of variables.

The respondents displayed great variability in terms of age (22 to 64 years) and professional background (15 graduate students in various fields, 54 entrepreneurs with at least three years of experience, 27 real estate employees, and 29 unemployed or retirees). Of the 125 questionnaires collected, we eliminated 5 because their mean response to the question “Are you confident in your choices?” were at the low end of the scale. The final sample therefore consisted of 120 respondents (65 men; mean age of 35 years).

3.4. Model

The data analysis had two objectives. First, we need a modeling framework that permits testing of the hypotheses related to the main effects ($H_1$–$H_4$) and moderator/interaction effects ($H_5$–$H_8$). Second, we hope to obtain preference partworths at the individual level, so that we can quantify the direction and amplitude of the trade-offs that respondents are willing to make among different DPT dimensions. These trade-offs are of interest to firms that want to align their DPT strategy with the individual consumers’ preferences. Although ranking tasks provide a natural context for respondents to express their preferences, they pose a modeling challenge for estimating individual parameters. As an illustration, suppose that a respondent has a very strong preference for visible DPTs, as opposed to the not visible ones. Of the six ranking tasks, three integrate the visibility dimension (visibility–justification, visibility–imposition, and visibility–surprise), for a total of six DPTs labeled visible (visible-justification, visible-imposition, visible-surprise, etc.). If the respondent has a very strong preference for visible DPTs and ranks them systematically as top choices—such that visible-something DPTs are systematically preferred to not-visible-something DPTs—then this respondent’s individual likelihood function remains undefined. It asymptotes to its maximum value as the parameter estimate for visibility goes to infinity.

To circumvent this challenge, we analyzed the conjoint data using a hierarchical Bayesian framework with individual random effects (e.g., Lenk, DeSarbo, Green, & Young, 1996). This framework provides a natural structure to express the customers’ preferences for DPT attributes as a function of their personality traits; the individual random effects also allow each respondent to deviate from average group preferences while preventing the individual parameter estimates from stretching to infinity. We used the following notations:

\[ i \text{ Respondents 1 to N, with } N = 120. \]
\[ a \text{ Attributes (1 = justification, 2 = imposition, 3 = visibility, 4 = surprise).} \]
\[ t \text{ Traits (1 = need for distinction, 2 = negotiation proneness).} \]
\[ p \text{ Profiles (DPT) presented on each ranking task, from 1 to } P, \text{ where } P = 4. \]

To model the ranking data among $P$ available profiles, we can write the likelihood as a sequence of $P - 1$ consecutive choices, often referred to as exploded logit/probit models (Begg, Cardell, & Hausman, 1981; Combes, Laurent, & Michael, 2008). The respondent first selects the most preferred option out of the $P$ profiles, then picks the second most preferred option out of the $P - 1$ remaining profiles (most preferred option excluded), and so on, until there is only one option left. Imagine, for example, that a respondent ranks four options, A, B, C, and D, by identifying A as the most preferred option (rank 1), C as the second most preferred (rank 2), and then D and B (ranks 3 and 4). Noting that $\pi(A|ABCD)$ is the estimated probability that option A will be the preferred option among the four, the full likelihood function for this consumer is as follows:

\[ L = \pi(A|ABCD) \cdot \pi(C|BCD) \cdot \pi(D|BD). \]
We model the probability of the choice among $P$ options using a logit function:

$$
\pi_{ip} = \frac{e^{\mu_{ip}}}{\sum_{p=1}^{P} e^{\mu_{ip}}}
$$

(2)

and

$$
\mu_{ip} = \sum_{a=1}^{4} \phi_{ia} X_{ap}
$$

(3)

where

$\pi_{ip}$ Probability that respondent $i$ chooses profile $p$ out of $P$ profiles

$\mu_{ip}$ Respondent $i$’s preference for profile $p$

$X_{ap}$ Attribute $a$ of profile $p$

$\phi_{ia}$ Preference of respondent $i$ for attribute $a$; because the model is fit on ranking data, there is no need for an intercept in the model.

The probability that respondent $i$ first selects profile $p$ out of the $P$ available profiles is a function of the respondent’s preference for that profile, compared with the respondent’s preferences for all available profiles. These preferences are expressed as a linear combination of preference partworths for each attribute describing the profiles.

The next level of the hierarchy expresses individual preference partworths in terms of personality traits, plus an individual random effect that captures residual preferences that are not explained by traits. That is,

$$
\phi_{ia} = T_{ia} + \epsilon_{ia}
$$

(4)

where

$\phi_{ia}$ Preference of respondent $i$ for attribute $a$

$T_{ia}$ Trait of respondent $i$, where $T_{ia} = 1$ by convention (intercept)

$\epsilon_{ia}$ Random component of respondent $i$ on attribute $a$, where $\epsilon_{ia} \sim N(0, \Sigma)$ and $\Phi$ and $\Sigma$ are empirically estimated from the data (empirical Bayes estimation with no hyperprior structure)

Before estimating the model, we mean centered and standardized the individual traits $T_{i}$ so that the intercepts $\phi_{ia}$ represent the average preference partworths for each attribute at the population level (to test $H_{2}$–$H_{4}$). The values of $\phi_{i1}$ and $\phi_{i2}$ serve to test the moderating influences of the need for distinction and negotiation proneness on the individual preferences ($H_{2}$–$H_{4}$). Finally, we estimated the model using an empirical Bayesian approach and the Metropolis–Hastings algorithm. The burn-in period was 10,000 draws, with the marginal posteriors estimated on 20,000 subsequent draws. The convergence, mixing, and rejection rates were satisfactory.

3.5. Results

In Table 3, we report the posterior means and the 95% confidence intervals of the $\phi$ parameters for this study; Fig. 1 reveals the distributions of $\phi_{ia}$ (Eq. (4)), which represent the preferences for the four dimensions of DPT estimated at the individual level. These preferences encompass the intercepts (average preferences at the population level), the moderating influences of the need for distinction and negotiation proneness, and individual random effects (i.e., individual variability not captured by the former variables). The heterogeneity in the preference partworths is evident, which highlights caveats for firms that anticipate that they can provide DPT without accounting for individual preferences.

At the average population level, the respondents preferred justified (.476, $p < .01$) and surprising (.472, $p < .01$) DPTs in support of $H_{1}$ and $H_{2}$. However, contrary to $H_{2}$, they also preferred imposing DPTs (.124, $p < .01$). Although unexpected, this effect can be explained in hindsight by the utilitarian facet of DPT. In the imposing condition in Study 1, the respondents imagined that a table had been booked for them in a full restaurant, where several other people were waiting. This preferential treatment provided a true utilitarian benefit: avoiding a long wait. In contrast, in the non-imposing condition, the restaurant was not full so that all customers, privileged or not, could sit down immediately, and the preferential treatment provided no utilitarian benefit. These expressed preferences suggest that the utilitarian facet of DPT prevails over its imposition facet (which we examine in Study 2).

The visibility parameter was not significantly different from 0 (−.027, $p = .30$), so we must reject $H_{3}$. However, the result does not indicate that the respondents were indifferent to this dimension. Quite the contrary, and consistent with De Wulf et al.’s (2001) assertion that some people are embarrassed by being openly favored, whereas others feel delighted by the attention, we illustrate in Fig. 1 that the respondents varied wildly in their preferences for visibility.

In terms of moderating effects, the respondents with a high need for distinction expressed a stronger preference for imposing (.408, $p < .01$) and visible (.888, $p < .01$) DPTs, in strong support for $H_{1}$ and $H_{2}$. These respondents also preferred the surprising DPT (.208, $p < .01$), which we did not hypothesize, although it makes sense in light of the increasing number of preferential treatments that modern consumers receive. As many researchers and practitioners note (e.g., Kumar & Shah, 2004), many consumers enroll in multiple loyalty programs in a single industry (e.g., loyalty cards from several grocery stores). This inflation has shaped a culture of entitlement, in which consumers are accustomed to receiving preferential treatment and come to expect it (Boyd & Helms, 2005). In this context, surprising DPTs are rare, which strengthens their distinguishing power and flattens the distinction seekers even more.

People with high scores on the negotiation proneness scale evaluated justified (−.374, $p = .04$) and surprising (−.288, $p < .01$) DPTs less favorably, confirming $H_{3}$ and $H_{6}$. These respondents like to take the initiative in the negotiation process, and they enjoy it more when they obtain unjustified preferential treatments. The interaction effect between negotiation proneness and visibility is also positive and significant (.173, $p < .01$). This effect was not hypothesized, though in hindsight it appears natural: what is the fun in demonstrating one’s negotiation skills if they go unnoticed?

In total, we confirm six of our eight hypotheses ($H_{1}$–$H_{6}$). The results for one hypothesis do not reach significance ($H_{4}$), and one parameter estimate is significant but opposite to the direction that we hypothesized ($H_{2}$).

4. Study 2: specialty store

4.1. Study objectives

Demonstrating heterogeneity in consumer preferences is valuable because it shows the urgent need to fine-tune DPT to customers’ expectations, but it provides little practical guidance to companies. It might be tempting to offer DPTs that please the majority of customers, but our results suggest that such an undifferentiated approach would be dangerous. The majority of respondents in Study 1 preferred justified, imposing, visible, and surprising DPTs,
but these preferences correspond to the exact preferences of only 28 respondents, or 23% of the sample. For the remaining 77%, some customers would feel embarrassed by the visible attention, guilty about imposing on other customers, or frustrated that they did not initiate the negotiation themselves.

Although we have shown that heterogeneity is partly explained by psychological traits such as the need for distinction or negotiation proneness, such moderators are not observable and are of little use to firms. Frontline employees are often in charge of selecting the customers who will receive DPT, as well as the characteristics of the DPT that they offer. They may have no information about the customers in front of them, yet they need to make a decision on the spot. To provide practical guidelines, it is therefore useful to explore whether preferences vary significantly in accordance with characteristics that are readily observable by frontline employees, such as age and gender.

Accordingly, the goals of Study 2 are threefold: (1) to validate the robustness of the Study 1 findings in a different managerial context; (2) to disentangle the respective roles of the imposition and utilitarian facets of DPT, and (3) to measure whether observable characteristics, such as gender and age, can guide the frontline employees in tailoring DPT.

4.2. Experimental setting

In Study 2, we asked another group of adult respondents to imagine a hypothetical scenario: they were shopping for a computer at a specialty store, and when they reached the checkout line, the salesclerk allowed them to skip ahead. The survey was similar to that for Study 1, with two adjustments. First, we manipulated imposition but kept the utilitarian benefit of the DPT constant, such that both high and low imposition conditions provided the same utilitarian benefit, namely, customers gained the same amount of time (see Table 2). Second, we adopted a slightly shorter scale for the need for distinction (two items instead of three) and negotiation proneness (two items instead of four).

4.3. Sample

The sample consisted of 110 adult respondents, 58 men (53%) and 52 women, all between the ages of 22 and 55 years (average = 36). They were participating in a medical seminar that took place in a city different from the location in Study 1. However, the data collection followed the same procedure; the respondents completed the six matrices then, three days later, they filled in the NPT scale.

4.4. Results

We used the model and estimation procedure from Study 1 and report the results in Table 3. Despite the new context, the results were highly consistent. On average, the respondents preferred justified (.193, p < .01) and surprising (.043, p < .01) DPTs, in support of H1 and H4. As in Study 1, the preferences for visibility indicated so much variance that the hypothesized main effect did not achieve significance (.038, p = .34), so we rejected H2. Also consistent with our previous findings, in contrast with H3, imposition had a significant, positive impact on the respondents’ preferences (.386, p < .01). Even when the utilitarian benefit of the DPT remained constant, a DPT that imposed on others had a greater perceived value, though the amplitude of the effect was smaller in this study (.386 vs. 1.246).

In terms of the interaction effects, the results confirmed all of our hypotheses. The higher the consumer’s need for distinction, the more

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Simple model</th>
<th>Complete model with interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main effect</td>
<td>Need for distinction</td>
</tr>
<tr>
<td>Study 1 (table at a hotel restaurant)</td>
<td>Justification</td>
<td>.600*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.506, .705]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imposition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.1030, 1.272]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.−211, −.079]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surprise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.231, .558]</td>
</tr>
<tr>
<td>Study 2 (skip waiting line in store)</td>
<td>Justification</td>
<td>.193*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.157, .243]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imposition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.195, .288]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.123, .200]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surprise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.008, .124]</td>
</tr>
</tbody>
</table>

Notes: The parameter estimates that are significant at p = .05 are in bold.

Fig. 1. Distribution of individual preference parameters for the four dimensions of DPT, Study 1. Notes: Of the 120 respondents, 25 prefer non-surprising DPTs, 73 prefer surprising DPTs, and 22 have an individual posterior mean that is not significantly different from 0 (the 95% interval of their posterior distribution contains 0).
he or she preferred imposing (.948, p < .01) and visible (1.101, p < .01) over non-imposing and non-visible DPTs, in support of H3 and H6. Again, we found support for a non-hypothesized, positive, significant effect of surprise (.659, p < .01) on the distinction seekers’ preferences. Furthermore, people prone to negotiate experienced greater satisfaction when the DPT that they received was not justified (−1.321, p < .01), in support of H2; they also preferred to stay in control of the negotiation process and disliked surprising DPTs (−.964, p < .01), in support of H8. Thus, despite the different context and sample, the results remained quite robust across Studies 1 and 2.

4.5. Managerial guidelines

We split the Study 2 sample by gender (men vs. women) and age (21–35 years vs. 36–65 years, corresponding to a median split), then performed two-tailed t-tests to check for significant differences in the psychological traits and preferences for justification, imposition, visibility, and surprise (Table 4).

In terms of the psychological traits, we found that the younger men scored twice as high as the older women on the negotiation proneness scale (4.84 vs. 2.83). The younger women scored the highest on distinction seeking (5.10), a trait that tended to decrease with age (from 4.59 to 3.40 for men and from 5.10 to 3.95 for women).

While justified, imposing, visible, and surprising DPTs earned the highest utility on average (i.e., the base DPT), the results in Table 1 show that significant improvements can be achieved by tailoring the DPT to various demographic groups. Older men tend to strongly dislike visible DPTs (consistent with their low score on the distinction seeking scale), so making sure that the firm does not grant a DPT to them ostentatiously and in the view of all other customers should increase their utility eightfold compared with the base DPT. The exact opposite result emerged for younger women (highest on the distinction seeking scale), for whom visibility is an essential characteristic of a delightful DPT. Younger men, the most prone to negotiate, tend to appreciate being in charge and benefitting from unjustified DPTs; ensuring that the granted DPTs are unjustified and unsurprisingly tends to increase their utility ten-fold. In all, these results provide interesting hints for how sales representatives should tailor the DPTs to the profiles of the customers that they encounter. Table 4 provides the characteristics of the ideal DPT for each demographic group.

### Table 4

Average preference partworths for DPT dimensions for the population as a whole and for different age and gender groups.

<table>
<thead>
<tr>
<th>Distinction seeking</th>
<th>Population as a whole</th>
<th>&lt;36 years</th>
<th>≥36 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>Justified</td>
<td>4.23</td>
<td>4.59</td>
<td>5.10a</td>
</tr>
<tr>
<td>Imposing</td>
<td>3.71</td>
<td>4.84c</td>
<td>3.56</td>
</tr>
<tr>
<td>Justification</td>
<td>4.33</td>
<td>−3.93c</td>
<td>5.73</td>
</tr>
<tr>
<td>Imposition</td>
<td>4.41</td>
<td>5.75</td>
<td>8.69</td>
</tr>
<tr>
<td>Visibility</td>
<td>0.86</td>
<td>1.56</td>
<td>9.45b</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.46</td>
<td>−1.99b</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ideal DPT</th>
<th>Justified</th>
<th>Unjustified</th>
<th>Justified</th>
<th>Unjustified</th>
<th>Justified</th>
<th>Unjustified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Justified</td>
<td>Imposing</td>
<td>Visible</td>
<td>Unsurprising</td>
<td>Justified</td>
<td>Imposing</td>
</tr>
<tr>
<td></td>
<td>Unsurprising</td>
<td>Surprising</td>
<td></td>
<td></td>
<td>Unsurprising</td>
<td>Surprising</td>
</tr>
<tr>
<td>Average utility of a justified, imposing, visible and surprising DPT to that group</td>
<td>12.06</td>
<td>1.39</td>
<td>29.91</td>
<td>2.14</td>
<td>14.26</td>
<td></td>
</tr>
<tr>
<td>Average utility of DPTs tailored to the preferences of that specific group</td>
<td>20.00</td>
<td>13.23</td>
<td>29.91</td>
<td>18.66</td>
<td>17.40</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>+66%</td>
<td>+852%</td>
<td>+0%</td>
<td>+772%</td>
<td>+22%</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Justified, imposing, visible, and surprising DPTs tend to be favored by the population as a whole, but younger male respondents tend to prefer unjustified and unsurprising DPTs (consistent with their higher-than-average negotiation proneness). Tailoring the justification and surprise of the DPT increases their average utility from 1.39 to 13.23, an almost tenfold increase. Two-tail t-test: significant differences in bold at (a) p < .01 and (b) p < .05.

5. Discussion

5.1. Contributions

With their resource limitations, companies need to devote more resources to their most profitable customers rather than their least profitable ones (Bolton et al., 2004). Although the question of which customers should be privileged has received considerable attention (e.g., Venkatesan & Kumar, 2004), the question of how to do so efficiently had not been addressed. We focus on an understudied type of preferential treatment, namely, discretionary preferential treatments (DPT), and highlight the perils that companies may encounter should they ignore the heterogeneity in customers’ evaluations of these treatments.

Consumers evaluate DPTs along four dimensions (justification, imposition, visibility, and surprise). Despite their heterogeneity, in general consumers prefer DPTs that are justified, impose on others, and surprise the recipient. The customers’ need for distinction and negotiation proneness moderate these preferences, such that distinction seekers have stronger preferences for visible and imposing DPTs, whereas negotiators prefer DPTs that are neither justified nor surprising. Building on social comparison processes (Festinger, 1954), the distinction seekers’ preferences for visible and imposing DPTs reflect a need to perform public, downward social comparisons (Besser & Zeigler-Hill, 2010). The negotiators’ preferences for unjustified, non-surprising DPTs reflect a need to attribute downward social comparisons to themselves (negotiation skills), which increases the ego-inflating effects of the DPT (Major et al., 1991).

From a theoretical point of view, this research extends social comparison theory (Festinger, 1954) to a consumption context, where the consumers’ abilities to gain additional benefits enhance their feelings of self-worth. This research also introduces nuance to equity theory (Adams, 1963, 1965) and its extension into the principle of reciprocity. Granting benefits commensurate with each customer’s contribution does not necessarily enhance consumer satisfaction. Negotiators are particularly sensitive to the asymmetry of their relationships (they prefer unjustified DPTs). Moreover, by demonstrating that consumers differentially value the four dimensions of preferential treatment, this research elucidates the contradictory positions in prior research and practice (De Wulf et al., 2001; Hennig-Thurai et al., 2002). Not all consumers like the same preferential treatments, so marketers must exercise care when determining the preferential treatment that they will grant.
From a practical point of view, our results specify that companies should implement a differential approach to preferential treatment to align the DPTs that they grant with their individual customers’ preferences. This approach starts with the identification of distinction seekers and negotiators. In contexts in which they can identify and interact with customers (e.g., hotels, leisure industries), customer contact employees should infer and track customers’ need for distinction and negotiation proneness—just as preferences for pillow softness or favorite coffee brands are tracked in some customer relationship management systems. If no prior identification is possible (e.g., first-time visitor to a store), the identification must rely on more easily accessible variables. As we show, both gender and age, though clearly imperfect, offer some predictive power in terms of identifying distinction seekers and negotiators. Young women are overrepresented among distinction seekers, and they prefer visible, imposing, and surprising DPTs. Young men are overrepresented among negotiators, and they prefer to maintain control of the negotiation process. Older men tend to prefer not visible DPTs. Such basic demographics offer frontline employees some hints about the types of consumer that they are addressing. In turn, these employees can adapt their reward processes to the personality type and preferences of each customer they meet, which reduces the potential stress that they may encounter in their work role (Whiting, Donthu, & Baker, 2011).

5.2. Further research

This article paves the way for further explorations of DPTs. We list a few directions that we consider to be of particular interest. First, most customers prefer imposing to non-imposing DPTs, which violates one of our core hypotheses and contradicts Adams’s (1963, 1965) equity theory. This finding holds even when we keep the utilitarian benefits constant (Study 2); it is particularly worthy of further investigation. We suspect that beyond the common correlation between imposition and utilitarian value (as in Study 1), imposing DPTs carry a higher symbolic value: the firm demonstrates its willingness to sacrifice its other customers’ well-being to delight the recipient. Although imposition might cause some negative emotions, it also triggers a strong positive feeling of being special and valued, and we suspect that the latter effect might be stronger than the former. Additional research should disentangle these two effects.

Second, although our results appear robust to various contexts (a hotel restaurant in Study 1 and a retail store in Study 2), future research should investigate the impact of other situational variables, such as the presence or absence of significant others and the degree of distinctiveness for the privileges. Imagine, for example, being awarded a seat in business class while flying from London to Singapore. Traveling by yourself on a business trip, you might enjoy the favor. If you travel with your family and friends, though, being the only one to receive this DPT is embarrassing. It would therefore be interesting to study the moderating impacts of social context, as well as the degree of DPT distinctiveness, on consumers’ evaluations. Another notable situational variable is the cultural context. Because DPT creates unequal situations, and cultures vary greatly in the extent to which they expect and tolerate such inequalities (Hofstede, 1980), it may be that a country’s power distance score influences the patterns of DPT preferences. Uncertainty avoidance, or the extent to which members of a culture feel uncomfortable in unstructured situations (Hofstede, 1980), might be another interesting cultural dimension. Cultures with high uncertainty avoidance try to minimize unstructured situations with strict laws and rules, which provides a particularly challenging setting for negotiators and may increase their propensity to negotiate preferential treatments.

Third, we have limited our studies to the consumer’s perspective and investigated the impact of perceived justification, imposition, visibility, and surprise. Yet justification from the customer’s point of view may appear totally unjustified to the company. Similarly, the impact of a preferential treatment on non-privileged consumers likely appears different, depending on the perspective taken. For example, if a privileged consumer receives direct access to after-sale services, he or she might not be aware that the dedicated service provision excludes non-privileged others, who then must wait far longer for their service appointments. To increase the practical implications of the findings, additional research should work to reconcile consumer and company perspectives.

Fourth, non-privileged consumers’ perspectives on DPTs should also be investigated. Managerial decisions must balance the DPT effectiveness among favored customers against the impact on the wider audience. To delight one customer might not be wise if it upsets ten others. It also would be helpful to determine the combined impact of justification, imposition, and visibility on spectators; we anticipate that visible, imposing, unjustified DPTs for a third party might be more upsetting than non-visible, non-imposing, justified DPTs.

Fifth, because the first step of any rewarding process consists of identifying which consumers to privilege, additional research should specify how to select consumers in a non-contractual setting. Substantial research has established the relevance of customer lifetime value as a segmentation tool (e.g., Rust, Kumar, & Venkatesan, 2011), yet to date, no studies explicitly address the question of how to identify such top-tier consumers without historical data. Answering this question is of great importance to determine the efficient uses of DPTs.

We thus identify DPTs as a double-edged sword: on the one hand, they offer a rare opportunity to delight customers who feel more entitled, yet also more undistinguished; on the other hand, if improperly used, they can backfire, embarrass, and frustrate. We have elucidated this important customer relationship tool by specifying (1) the dimensions along which DPTs are evaluated, (2) the heterogeneity of customers’ preferences along these dimensions, (3) how customers’ preferences are moderated by psychological traits, and (4) how sales representatives can anticipate customers’ likely preferences using simple sociodemographic variables. We hope that this research sparks further interest in this fruitful and managerially relevant topic.

Acknowledgments

The authors would like to thank Timothy B. Heath and Christian Pinson for their helpful comments on earlier versions of this article.

Appendix A. Factor structure of the need for distinction and negotiation proneness scales

<table>
<thead>
<tr>
<th>Items</th>
<th>Butori (2010) sample (n = 224)</th>
<th>Study 1 sample (n = 120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Distinction</td>
<td>Negotiation</td>
</tr>
<tr>
<td>I generally do not like to be considered the same as any other customer</td>
<td>.887</td>
<td>.931</td>
</tr>
<tr>
<td>I do not like to feel like any customer</td>
<td>.814</td>
<td>.772</td>
</tr>
<tr>
<td>In general, I like to be treated differently from other customers</td>
<td>.798</td>
<td>.850</td>
</tr>
<tr>
<td>I need to feel that I am a customer who is granted special attention</td>
<td>.843</td>
<td>.840</td>
</tr>
<tr>
<td>Just for the fun of it, I often ask salespeople for special offers</td>
<td>.891</td>
<td>.907</td>
</tr>
<tr>
<td>I enjoy asking salespeople for special treatment</td>
<td>.854</td>
<td>.877</td>
</tr>
<tr>
<td>I enjoy negotiating advantages as much as actually using them</td>
<td>.817</td>
<td>.856</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>.865</td>
<td>.822</td>
</tr>
</tbody>
</table>
The impact of national brand introductions on hard-discounter image and share-of-wallet

Carlos J.S. Lourenço a,⁎, Els Gijsbrechts b,1

a Rotterdam School of Management, Erasmus University of Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
b Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

Hard-discounters (HDs) such as Aldi and Lidl are increasingly introducing national brands (NBs) into their private label (PL) dominated assortments. While there is evidence that this enhances sales in the categories where such NBs are added, little is known about how it affects consumers' overall perceptions of the HD and consequently the share of the customers' wallet. Using a unique data set that combines longitudinal information on a HD's perceptions, with that chain's assortment composition, we investigate the impact of NB introductions on the chain's overall value and assortment image, and spending share.

We show that introductions of NBs, in particular category leaders, may significantly contribute to a more favorable perception of the HD store. For positive price-value effects to materialize, HDs must offer these NBs at low-enough prices to maintain a reasonable price gap with the current private label offer. For the NB entry to enhance the HD's assortment perception, it must come with a sufficiently deep product line. However, there are limits to this approach. Introductions gradually lose effect as the share of NBs at the HD goes up. More importantly, ill-selected NB additions may backfire on the HD. Listing NBs that are not category-leaders, at prices too far above its private labels, deteriorates the HD's favorable value positioning — cutting into its core competitive advantage, and leading to notable reductions in share-of-wallet. We discuss the academic and managerial implications of these findings.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Sales of the top-10 discounters in the world are expected to grow by 50% from 2010 to 2015 (Planet Retail, 2010). By then, the German-based retailers Aldi and Lidl, pioneers of the hard-discount concept and ranking number one and two on the top-10 discounter chart, are both expected to hit the $100 billion mark (Planet Retail, 2010). This remarkable success of hard-discount retailers is rooted in their ability to practice low prices — 15 to 20% lower than those of large discounters like Wal-Mart (Wall Street Journal, 2009). Compared to traditional retailers, hard-discount stores (HDs) focus on own brands, and use a simplified ‘no-frills’ store format with little promotional activity — strategic options that translate into cost efficiencies in the supply-chain. Not surprisingly, these HDs have acquired a substantial share of grocery sales at the expense of mainstream retailers in Western Europe, and are growing consistently in the U.S. (Cleeren, Verbouwen, Dekimpe & Gielen, 2009; Planet Retail, 2010; Steenkamp & Kumar, 2009).

Nonetheless, hard-discount chains have realized that growth strategies based on prices are not without limits, and that an overreliance on price-based competition makes them vulnerable to incoming discounters. Partly due to this realization, HDs are increasingly introducing national brands (NB) into their merchandise offer. At Lidl, 30% of the assortment is now composed of NBs, roughly the same percentage as their contribution to total sales (Steenkamp & Kumar, 2009). Aldi, which not long ago had only private labels (PL) in its assortment, now lists well-known manufacturer brands in several product categories and markets, such as PepsiCo’s Quaker Oats, Kraft’s Oscar Mayer hot-dogs or Dole’s fresh fruit in the U.S., or Ferrero’s Nutella and Danone’s yogurts in Germany and eastern Europe — to name just a few. Even in countries where it stuck to the ‘strict PL’ policy to date, Aldi now announces the introduction of major NBs into its assortment (Distrifood, 2012) and “in a major break from tradition, [Aldi] is preparing to introduce brands across its entire global network” (Research Farm, 2012). For manufacturers, presence on the HDs’ shelves is a way to alleviate their dependency on mainstream retailers, who have extensively developed their own private label lines (Ailawadi, Pauwels, & Steenkamp, 2008) and put increased pressure on national brand margins to compete with the discounters’ success (Bloom & Perry, 2001).

Despite the strategic importance of NB introductions for both manufacturers and retailers (ter Braak, Deleersnyder, Geyskens & Dekimpe, 2013), little is known about their impact at hard-discount
stores. In an interesting study, Deleersnyder, Dekimpe, Steenkamp and Koll (2007) show that such introductions may entail a win–win situation: NBs grow their share relative to competing brands and hard-discounters gain share in the product category. However, the question remains how NB introductions – so directly at odds with the core positioning of the hard-discounter format – affect consumers’ overall perceptions of these chains, and, ultimately, the chains’ share-of-wallet (SOW). As research analysts put it: will brands become a major loyalty driver and tie in shoppers that are currently trying out the discounter? (Research Farm, 2012). If carrying more NBs changes the hard-discounters’ value positioning or assortment image, NBs’ impact is likely to extend to other categories in the store (where no NBs are offered). Specifically, to the extent that increased presence of NBs (at prices above the HD’s typical PL offer) jeopardizes the HD’s reputation of providing excellent value-for-money across-the-board, this may deteriorate overall store performance. Indeed, consumer perceptions are shown to be key drivers of store choice and spending (e.g. Cox & Cox, 1990; Srivastava & Lurie, 2004; van Heerde, Gijsbrechts, & Pauwels, 2008), which seems to be particularly true for hard-discounter shoppers (Lourenço & Gijsbrechts, 2010). If, as advocated in a recent Nielsen report, “Good Value is a Matter of the Mind” (Nielsen, 2008, p.6), then HDs have an interest in tracing the effect of higher NB presence on their assortment and value perceptions – making sure their discount positioning does not blur (Deleersnyder et al., 2007).

This paper sets out to address these issues. In so doing, we add to the work of Deleersnyder et al. (2007), who are the first to study NB introductions and the characteristics that make for their success at hard-discounters. Our research complements theirs, in that we take a different perspective and bring different outcome metrics to the table. Specifically, while Deleersnyder et al. (2007) identify NB listings that meet the ‘joint’ interests of the manufacturer and the retailer and consider brand- and category sales as the focal variables, we take the perspective of the hard-discounter and consider implications for the chain as a whole in terms of image and share of wallet. As such, our work fits into the recent call of Srinivasan, Vanhuele and Pauwels (2010) to study the effect of marketing actions on mindset metrics and subsequent ‘hard’ performance measures. Not only does the inclusion of mindset metrics enhance the explanatory power of response models, it may also lead to a richer understanding and more actionable managerial recommendations (Srinivasan et al., 2010).

We focus on the following research questions. First, how do national brand introductions in HD stores affect the assortment and value positioning of these stores? Second, do these effects depend on the characteristics of NBs and of the category where they are introduced? Third, what are the effects of these NB additions, and their ensuing image consequences, on the HD’s share-of-wallet? We empirically test for the presence and size of these effects using a unique data set for the Belgian market, obtained from GfK. The data set combines (i) information on the HD chain’s assortment composition over time (i.e. NB introductions and deletions), with (ii) longitudinal information on its assortment and value perceptions among individual households, along with (iii) those same households’ weekly purchases at the HD and competing chains.

The remainder of the paper is organized as follows. In the next section, we briefly review the relevant literature and develop the conceptual framework. Section 3 describes the methodology, data, and variable operationalizations. The empirical results are presented in Section 4. Section 5 discusses the findings, implications, and limitations, along with suggestions for future research.

2. Conceptual framework

2.1. Background

Store image plays a pivotal role in retailers’ strategies (Ailawadi & Keller, 2004; Steenkamp & Wedel, 1991), and has been shown to strongly influence consumer store choice and spending (e.g. Lourenço & Gijsbrechts, 2010; Nielsen, 2008; van Heerde et al., 2008). Defined as the way the store is perceived in the shopper’s mind (Martineau, 1958), store image is typically seen as a multidimensional construct, with price, quality and variety of the assortment as its most important dimensions (Hildebrandt, 1988; Mazursky & Jacoby, 1986). The formation of a retailer’s image is a dynamic process (Büyükkurt, 1986; Mazursky & Jacoby, 1986; Nyström, Tamsons, & Thams, 1975) – perceptions being updated as new information comes in. Given the complexity of stores’ offers, consumers have incomplete information and are uncertain about retail stores, thus resorting to available perceptual cues when inferring or updating retailers’ overall positioning (Feichtinger, Luhmer, & Sorger, 1988; Mägi & Julander, 2005).

In consumer packaged goods, manufacturer brands constitute important perceptual cues (Dawar & Parker, 1994; Grewal, Krishnan, Baker & Borin, 1998) that contribute to the image of the retailer carrying them (Ailawadi & Keller, 2004). Clearly, brand presence should affect consumers’ perceptions of a store’s assortment. Also, brands are usually deemed informative about quality, as quality is often perceived with uncertainty (Richardson, Dick, & Jain, 1994; Zeithaml, 1988). Moreover, in a store context, where price information is complex and ambiguous (Hamilton & Chernev, 2010a), consumers may be uncertain about a store’s overall expenses too (Alba, Broniarczyk, Shimp & Urbany, 1994), and the presence of known brands may prove helpful in the formation of price and ‘value-for-money’ perceptions (Biswas, Wilson, & Licata, 1993; Monroe, Grewal, & Compeau, 1991). We build on these insights below.

2.2. Conceptual framework

Fig. 1 summarizes our framework on how the introduction of NBs influences the hard-discounter’s perceptions and, subsequently, share-of-wallet.

2.2.1. Impact on value image

NBs are typically (still) perceived to have higher quality than private labels (Kumar & Steenkamp, 2007; Steenkamp, van Heerde & Geyskens, 2010). Being such ‘beacons of quality’, they are likely to enhance the perceived quality of the hard-discounter’s offer. At the same time, national brands are generally higher-priced than private labels (Ailawadi, Neslin, & Gedenk, 2001; Steenkamp et al., 2010). Hence, NB introductions may also lead consumers to infer that the HD has become more expensive (Hamilton & Chernev, 2010a; Mägi & Julander, 2005). The impact on the HDs ’value-for-money’ positioning – a key ingredient of its success – is thus not clear a priori, but depends on the net outcome of these two forces.

2.2.2. Impact on assortment image

Even if total assortment size remains unchanged, NB presence may enhance the consumer’s overall perception of assortment appeal and variety offered by the store (Oppewal & Koelemeijer, 2005). In contrast with the own labels carried by hard-discounters, NBs invest heavily in their ’aura of uniqueness’ (Geyskens, Gielens, & Gijsbrechts, 2010) and, therefore, have often built a segment of loyal customers. As shown by Broniarczyk, Hoyer and McAlister (1998), consumers’ variety perceptions are shaped by the presence of a “favorite brand”. Hence, carrying NBs is likely to contribute to the HDs overall assortment appeal – creating a more favorable assortment image.

2.2.3. Moderating effects

Clearly, not all NBs are alike, and we expect the strength of these effects to be moderated by the following brand- and category characteristics.

2.2.3.1. Price gap. The size of the perception adjustment may depend on the price charged for the NB at the HD store. Like Deleersnyder...
et al. (2007), we see a potential role for two distinct price gaps: (i) the "between-store price gap", i.e. the difference between the NB price at the HD compared to that of traditional stores, and (ii) the "within-store price gap", i.e. the deviation between the HD's price for the NB and its PL offer. The *between-store price gap* comes into play when shoppers engage in cross-store comparisons (Gauri, Sudhir & Talukdar, 2008); a larger gap (higher NB price at traditional stores relative to the HD) possibly enhancing the perception that the HD offers good 'value-for-money' (Hamilton & Chernev, 2010a). However, consumers may also build their price beliefs by focusing on cues inside the store (Hamilton & Chernev, 2010a; Lourenço & Gijsbrechts, 2010), and use the premium charged for the NB compared to the HD's PL to adjust the discounter's image. This within-store price comparison may trigger a dual effect. On the positive side, a large price gap may signal that the NB is in a different 'quality league', thereby lifting the quality perception of the HD's offer (Deleersnyder et al., 2007). Also, the price of the vertical extension may be 'contrasted' with that of the prevailing PL offer, emphasizing that the HD's PL offer is cheap (Hamilton & Chernev, 2010b). In such cases, a larger *within-store price gap* (NB more expensive relative to the PL) may trigger a more favorable value image. On the negative side, consumers may 'integrate' the new (high) NB price into their overall price beliefs. In that case, larger price gaps translate into a less favorable value image of the store (Hamilton & Chernev, 2010b). The net outcome depends on the strength of these opposing forces. Which one will prevail, we leave as an empirical issue.

2.2.3.2. Leading brands. Similar to Cielens (2012), we expect brand positioning to play a role. National brands that hold a leading position in their category are likely to be better known and trusted, and positively contribute to the HD's perceived quality. Hence, ceteris paribus (i.e. after controlling for price differences), we expect NBs that are category leaders to more favorably influence perceived value-for-money. Moreover, such leading brands are more likely to be consumers' 'favorites', and, as such, more apt to upgrade the HD's assortment perception and appeal (Broniarczyk et al., 1998; Sloot & Verhoef, 2008), which, again, suggests a positive moderation.

2.2.3.3. Brand line depth. A NB introduction is likely to more strongly enhance assortment perceptions as more stock-keeping units (SKUs) are introduced. Not only does this make the NB shelf presence more prominent, but variety within a brand line has also been shown to increase assortment appeal (Boatwright & Nunes, 2001; Borle, Boatwright, Kadane, Nunes & Shmueli, 2005). Hence, we expect a positive moderation of number of introduced NB-SKUs on the HD's assortment image.

2.2.3.4. Category purchase frequency. NB introductions may also be more influential in frequently purchased categories. Changes in such categories are more likely to be noticed (Hamilton & Chernev, 2010a), and may matter more to consumers. Moreover, product categories that are more often purchased have greater potential to enhance the (perceived) utility of shopping at the store (Bell, Ho & Tang, 1998; Inman, Winer, & Ferraro, 2009).

2.2.3.5. Cumulative share of NB SKUs. Finally, we expect the impact of NB introductions to depend on how many NB products the discounter already carries at the time of introduction. Previous studies already uncovered a nonlinear relationship between store patronage and brands (SKUs) offered (Borle et al., 2005; Briesch, Chintagunta & Fox, 2009), in which item 'uniqueness' plays an important role. Building on those insights, we expect that HDs where NBs are absent or where their presence is negligible, may experience momentum: new introductions being evaluated more positively as the share of NB products in the HD assortment increases. This leveraging may occur because consumers take better notice of the NB offer or begin to find it extensive enough to weigh on their perceptions. However, while this phenomenon may be observed initially, extra introductions may lead to little incremental

![Conceptual framework](image-url)
assortment appeal, and jeopardize the perception of low (PL) prices across-the-board. We thus expect the cumulative presence of NBs to have a moderating effect for value and assortment image, but leave the direction of the effect as an empirical issue.

2.2.4. Impact on share-of-wallet
By listing NBs, HDs aim to build stronger store loyalty and enhance shoppers’ share of spending at the store (Deleersnyder et al., 2007; Planet Retail, 2010). Given the critical role of images in consumer store choice and spending (e.g. Allawadi & Keller, 2004; Baker, Parasuraman, Grewal & Voss, 2002; van Heerde et al., 2008), any changes in the HD’s ‘good value-for-money’ reputation or perceived assortment variety from adding NBs, are, indeed, likely to affect the shoppers’ budget share spent at the store. Fig. 1 includes these effects, and portrays how NB listings, through their impact on the hard discounter’s reputation, translate into store share-of-wallet. Similarly, consumers may only partially update their store perceptions upon NB introduction, we incorporate lagged assortment and value images. Similarly, consumers’ previous share-of-spending at the HD chain enters as a baseline to mainly materialize through the value and assortment perceptions. Still, it is not excluded that they also affect purchases directly, without ‘leaving their footprint’ on the store’s overall image. For completeness, we allow for such direct effects in our framework, and empirically test for their presence, but without specifying a direction for the main effect, or postulating specific moderating effects.

Table 1 summarizes the expected NB introduction effects.

2.2.5. Controls
Fig. 1 captures these effects and includes other factors that we control for. We expect NB introduction effects to occur over and above actual changes in price and assortment size – which we therefore include as separate variables. Also, consumers may differ in their ‘baseline’ perceptions of HDs, something we control for by including ‘hard-discounter proneness’ as a household characteristic. To account for the fact that images are ‘sticky’ (Alba et al., 1994) and that consumers may only partially update their store perceptions upon NB entry, we incorporate lagged assortment and store images. Similarly, consumers’ previous share-of-spending at the HD chain enters as a control for its current SOW. Finally, HDs may also delist NBs from the shelves, and this too may influence consumers’ perceptions of and spending at the store.

Fig. 1 also indicates that store assortment and value perceptions may be correlated directly, over and above the effect of their joint drivers. Indeed, favorable attitudes towards a store may result in a “halo” effect (Holbrook, 1983): consumers holding a good value image of the HD chain associating it with a favorable assortment image too. Our methodological approaches include such spillovers.

3. Method and data
3.1. Method
To address our research questions, we proceed in two steps. First, we assess the effect of NBs on households’ assortment and value perceptions of the HD, thereby accommodating moderating brand- and category characteristics. Next, we investigate the impact on the households’ SOW spent at the HD. We discuss both steps in turn.

3.1.1. Stage 1: Impact of NB introductions on HD image
To assess the effect of NB entries on the HD’s assortment and value image, we estimate the following system of equations (for an overview of the notation, see Table 2):

\[ \begin{align*}
\text{Value}_i^t &= \alpha_1 + \sum_j \alpha_2 \text{NB Intro}_j^t + \sum_j \alpha_3 \text{NB Exit}_j^t + \alpha_4 \text{HD Price}_e^t + \alpha_5 \text{HD Proneness}_h^t + \epsilon_i^t \\
\text{Assort}_i^t &= \beta_1 + \sum_j \beta_2 \text{NB Intro}_j^t + \sum_j \beta_3 \text{NB Exit}_j^t + \beta_4 \text{HD #SKU}_s^t + \beta_5 \text{HD Proneness}_h^t + \alpha_i^t
\end{align*} \]  

with:

\[ \begin{align*}
\alpha_1 &= \mu_1 + \mu_2 \text{Cat Freq}_i + \mu_3 \text{CUM NB Shares}_i + \mu_4 \text{Leader}_f + \\
&\quad + \mu_5 \text{PGap between}_i + \mu_6 \text{PGap within}_i \\
\beta_1 &= \lambda_1 + \lambda_2 \text{Cat Freq}_i + \lambda_3 \text{CUM NB Shares}_i + \lambda_4 \text{Leader}_f + \lambda_5 \text{LDepth}_d
\end{align*} \]

where \( i \) (\( j \)) refers to a specific NB entry (exit) event, \( \text{Value}_i^t \) and \( \text{Assort}_i^t \) denote consumer \( h \)'s overall value and assortment image of the HD store in period \( t \), respectively, and \( \epsilon_i^t \) and \( \omega_i^t \) are normally distributed error terms, with zero mean and variances \( \sigma^2_{\epsilon} \) and \( \sigma^2_{\omega} \) respectively. By letting \( \text{cov}(\epsilon_i^t, \omega_i^t) = \rho \sigma_{\epsilon} \sigma_{\omega} \), we account for a possible correlation between a household’s assortment and value scores at a given point in time. Also, to accommodate unobserved household heterogeneity, we include household-specific random effects, which we assume to be normally distributed: \( \epsilon_i^t \sim N(0, \sigma^2_{\epsilon}) \) and \( \omega_i^t \sim N(0, \sigma^2_{\omega}) \).

Key to the model are the effects of NB introductions at the HD in the course of the observation period, which we break down into a main effect (captured by \( \text{NB Intro}_i \)), and a moderating influence of (1) the purchase frequency of the category in which the brand is introduced (\( \text{Cat Freq}_i \)), (2) the (cumulative) share of NB products already available in the HD assortment at the time of introduction (\( \text{CUM NB Shares}_i \)), (3) whether or not the brand is a category leader (\( \text{Leader}_f \)), (4) the NB’s price gap between the HD and the traditional chains (\( \text{PGap between}_i \)), in the value equation), (5) the NB–PL price difference within the HD store (\( \text{PGap within}_i \)), in the value equation), and (6) the depth of the NB line (\( \text{LDepth}_d \)), in the assortment equation.

As control variables, we include the HD’s overall price level \( \text{HD Price}_e \), (in the value equation) and assortment size \( \text{HD #SKU}_s \) (in the assortment equation), as well as the removal of NBs from the HD’s shelves (\( \text{NB Exit}_i \)). (Note that, since the number of exits is limited, we include these only as controls, and do not disentangle their effect.

As further explained below, because of differences in available data for these two types of dependent variables (half-yearly data for value and assortment image, weekly data for SOW), we estimate these models in two stages instead of parameterizing one overall model. As a post-hoc check, we correlated the aggregated residuals of the SOW equation, with the corresponding residuals of the image equations, and found the correlations to be very low (\( -.007 \) and \( -.07 \), for value and assortment image, respectively).

Table 1
<table>
<thead>
<tr>
<th>Summary of expected effects.</th>
<th>Expected effect</th>
</tr>
</thead>
</table>

| Impact on value image of NB Introduction (main effect) | (++) |
| Moderators | |
| Between-store price gap | (+) |
| Within-store price gap | (+) |
| Leader brand | (+) |
| Category purchase frequency | (+) |
| Cumulative NB SKU Share | (+/-) |

| Impact on assortment image of NB introduction (main effect) | (+) |
| Moderators | |
| Leader brand | (+) |
| Brand line depth | (+) |
| Category purchase frequency | (+) |
| Cumulative NB SKU share | (+/-) |

| Impact on share-of-wallet of NB introduction (main effect) | (+/-) |
| Value image | (+) |
| Assortment image | (+) |
by brand or category). As already mentioned, to further address heterogeneity in the households’ ‘base’ perceptions, we include a household’s general propensity to visit (any chain of) a hard-discounter format (HDProne) as an extra control. Last but not least, to capture the ‘stickiness’ of perceptions, we include lagged assortment and value scores (Assort, −1 and Value, −1, respectively).

Table 2
Notation overview.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value equation</th>
<th>Assortment equation</th>
<th>SOW equation</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value _t+1</td>
<td>√</td>
<td></td>
<td></td>
<td>Value (assortment) image of household h in period t: “How do you rate the (prices) (merchandise quality) (assortment) of products in this store (Lidl)?” on a scale from 1 (= least favorable) to 9 (= most favorable).</td>
</tr>
<tr>
<td>Assort _t+1</td>
<td>√</td>
<td></td>
<td></td>
<td>Share-of-wallet of household h in period t+1 as calculated: the amount spent at Lidl, divided by the total amount spent on grocery shopping in the course of that week.</td>
</tr>
<tr>
<td>SOW _t+1</td>
<td></td>
<td></td>
<td></td>
<td>1-week lagged value of SOW;</td>
</tr>
<tr>
<td>Values, i.t−1</td>
<td>√</td>
<td></td>
<td></td>
<td>Household h’s discounter trip share, i.e. number of household trips to discounters (Aldi and Lidl in our data set) divided by the total number of household trips.</td>
</tr>
<tr>
<td>Assort, i.t−1</td>
<td></td>
<td></td>
<td></td>
<td>Step dummy identifying introduction i (exit j) of a NB in time period t: NBs are brands that (i) are not flagged (by GKI) as PLs, and (ii) are sold at each of the three traditional supermarkets. Introduction (exit) weeks are weeks in which an SKU purchase of the NB was first (last) recorded at Lidl in the entire national panel.</td>
</tr>
<tr>
<td>SOW, i.t−1</td>
<td>√</td>
<td></td>
<td></td>
<td>(Cumulative) share of NB SKUs already available in the HD assortment at the time of introduction i.</td>
</tr>
<tr>
<td>HDProne _t</td>
<td>√</td>
<td></td>
<td></td>
<td>Purchase frequency of the category in which introduction i occurred: among category buyers (i.e. households with at least one purchase in the category), the yearly average number of times households purchase from the category.</td>
</tr>
<tr>
<td>NB _ Intro_h</td>
<td></td>
<td>√</td>
<td></td>
<td>‘Leader’ dummy variable for introduction i: 1 if the introduced NB is the top selling brand in the category market-wide, 0 otherwise.</td>
</tr>
<tr>
<td>Cat _ Freq_h</td>
<td></td>
<td>√</td>
<td></td>
<td>Line depth for introduction i: number of SKUs of the introduced NB at the time of introduction.</td>
</tr>
<tr>
<td>Leader_h</td>
<td></td>
<td>√</td>
<td></td>
<td>Between-store price gap for introduction i: log-transformed ratio of the (average of the) NB’s unit price at the top-three traditional chains, over its unit price at Lidl.</td>
</tr>
<tr>
<td>SKUst</td>
<td></td>
<td></td>
<td></td>
<td>Within-store price gap for introduction i: log-transformed ratio of the (average of the) NB’s unit price at Lidl, over the unit price of the corresponding PL at Lidl.</td>
</tr>
<tr>
<td>PGap _ between_h</td>
<td></td>
<td></td>
<td></td>
<td>Lidl’s (HD competitor’s (Aldi)) (top-3 mainstream supermarkets) actual overall price index in period t (Obtained by (1) dividing SKUs’ observed price per volume unit by the average unit price in the category across all weeks and stores; then (2) aggregating these across SKUs in the category, and (3) weighing the resulting category-level price indices by the category’s average (market-wide) share-of-wallet to obtain an overall price index of the store (see van Heerde et al., 2008)).</td>
</tr>
<tr>
<td>HD _ Price_h</td>
<td></td>
<td></td>
<td></td>
<td>Total number of different SKUs available at Lidl (HD competitor (Aldi)) (top-3 mainstream supermarkets) in week t, in thousands, expressed as a running monthly average.</td>
</tr>
</tbody>
</table>

For ease of interpretation, we mean-center the continuous brand- and category-moderators across NB entries, prior to calculating the interaction variables. This allows us to interpret the main effect as an ‘average’ introduction effect for these moderators. It also helps to see the intuition behind Eq. (1c): if a NB entry is below average on a certain moderator, a negative term is added in Eq. (1c), and the moderator variable in the form of a cumulative sum will decline.4

We estimate the main- and interaction parameters {τ0, α0, β0, σ2(α0, β0), α2, β2 to β5, μ1 to μ6 and λ1 to λ3} of the two image equation directly, using Seemingly Unrelated Regression (SUR) for unbalanced panels, with household-specific random intercepts (i.e. Stata’s xtsur).  

3.1.2. Stage 2: impact on HD share-of-wallet

Next, to assess the influence of NB introductions (and exits) on the HD’s share-of-wallet, we estimate the following model:

\[
SOW_{it} = \gamma_0 + \sum_{j} \gamma_j \text{NBIntro}_{ij} + \sum_{j} \gamma_j \text{NBExit}_{ij} + \gamma_1 \text{Value}_{ij} + \gamma_2 \text{Assort}_{ij} + \gamma_3 \text{SOW}_{ij-1} + \\
\gamma_4 \text{HDPrice}_{it} + \gamma_5 \text{HDCompPrice}_{it} + \gamma_6 \text{MainPrice}_{it} + \\
\gamma_7 \text{HDSKU}_{it} + \gamma_8 \text{HDCompSKU}_{it} + \gamma_9 \text{MainSKU}_{it} + \gamma_10 \text{It} 
\]

(2a)

with:

\[
\gamma_1' = \theta_1 + \theta_2 \text{CUM NBShare} + \theta_3 \text{Cat Freq} + \theta_4 \text{Leader} + \\
\theta_5 \text{PGap between} + \theta_6 \text{PGap within} + \theta_7 \text{LDepth},
\]

(2b)

4 We thank the (anonymous) AE for pointing this out.
where \( I^t \) are normally distributed error terms with zero mean and variance \( \sigma^2_\varepsilon \). Given that SOW is a relative measure, we include as controls not only the price and assortment size of the focal HD itself (\( HD_\text{Price}_i, HD_\text{Assort}_i \)), but also those of competing HDs (\( HD_\text{Comp}_\text{Price}_i, HD_\text{Comp}_\text{Assort}_i \)) and mainstream supermarket stores (\( Main_\text{Price}_i, Main_\text{Assort}_i \)), next to lagged shared-of-wallet at the focal HD.

As for the NB introductions, we allow them to affect SOW directly (through the main-effect step dummy NB \( INTRO_i \); and the moderating variables \( CUM\_NBShare_i, Cat\_Freq, Leader_i, LDepth_i, PCap\_between_i \), but also indirectly, through the value and assortment image variables ( \( Value_i, Assort_i \)). We accommodate household heterogeneity by using household-specific random effects: \( \gamma_0^N - N(\gamma_0, \sigma^2_{\gamma_0}) \). For estimation purposes, we again substitute Eq. (2b) into Eq. (2a), which, after rearranging terms, leads to the following equation

\[
\text{SOW}_i = \gamma_0 + \theta_1 \sum_{t-1}^{t} \text{NBIntro}_i + \theta_2 \sum_{t-1}^{t} \text{CUM\_NBShare}_i + \varepsilon_i
\]

and then estimate the parameters \( \{\gamma_0, \sigma^2_{\gamma_0}, \theta_1, \theta_2, \gamma_2\} \) directly.

### 3.2. Data and operationalizations

Our empirical analysis focuses on one of the most successful HDs in the world, the German-based Lidl, part of the Schwarz Group. A family-run business, Lidl opened its first store in 1973, following the concept created by its main competitor, Aldi. Lidl has been growing ever since and, in the last 15 years, opened new stores at the rate of one per day (Nielsen, 2008). Operating more than 10,000 stores in about 30 countries, Lidl reached close to $88 billion in sales in 2011 (Planet Retail, 2012). It is now the 2nd largest grocery retailer in Western Europe, and ranks 7th worldwide (Planet Retail, 2012).

We use data from Lidl’s operations in Belgium. With a consistent growth over the last years, Lidl is now the 5th grocery retailer in Belgium, just below Aldi (Planet Retail, 2012). Moreover, Belgium is one of the countries where NB presence at Lidl is an important recent phenomenon, such that it constitutes an excellent setting to study the effect of NB introductions at HDs — as outlined below.

#### 3.2.1. Operationalizations: dependent variables

##### 3.2.1.1. Store image data

Our data comprise information on consumers’ perceptions of store merchandise quality, price and assortment, obtained from members of GK’s national household panel. This panel includes over 4000 households that represent a stratified random sample of the Belgian population. The data cover a period of six years, from January 2005 to December 2010, and include 12 waves of surveys, each conducted during one week (around weeks 16 and 40 every year). In the course of that week, a random sub-sample of panelists is asked to judge the overall price level and other dimensions of visited stores. Specifically, upon returning home after a store visit and scanning their purchases along with their scanner-panel ID, panelists answer the questions:

1. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”
2. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”
3. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”
4. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”
5. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”
6. “How do you rate the (prices) (merchandise quality) (assortment) of products in this store?”

As such, panel members evaluate as many stores as actually visited during the survey week. In total, 1689 households rated Lidl in at least one survey wave. Survey responses (and store visits) are quite balanced over time (on average, each wave comprises ratings from 285 different households and a given household participates in two survey waves, not necessarily consecutively).

Fig. 2 shows the evolution of the image scores for Lidl, averaged across households. Three insights emerge. First, household perceptions about Lidl vary substantially in the course of our observation period (for instance, the average rating ranges from a low of 6.31 to a high of 7.66 for assortment, and from a high of 7.47 to a low of 7.05 for price — differences that are strongly significant: \( p < .01 \)). Second, closer inspection reveals that consumers’ price and quality ratings evolve similarly across survey waves (see Fig. 2), and, based on a factor analysis of the individual scores, represent a single construct capturing the store’s perceived ‘value-for-money’ (Eigenvalues: 1.547 and 1.453, first factor explaining 77% of the variance; Cronbach alpha = .706). The store’s assortment ratings, however, show a different over-time pattern (see Fig. 2) and cannot be merged reliably with either price (Cronbach alpha = .553 < .7) or merchandise quality ratings (Cronbach alpha = .636 < .7). Hence, and in line with our conceptualization, in the remainder of the analysis we use the summed ‘value’ scale — obtained as the summary of price and quality ratings — as one construct,\(^7\) while considering the effects on assortment image separately.\(^3\) Third, eyeballing the aggregate image evolutions, we note that some of them can be roughly linked to NB-related events. For instance, the upswing in waves 3 and 8 corresponds to periods with quite some NB introductions, whereas the dip in wave 7 coincides with few introductions and several exits. Still, the pattern is far from clear cut: some periods with many NB additions (e.g. wave 6) leading to opposite patterns for value and assortment image. Our analysis below sheds more light on the underlying forces.

---

4. The data also contained survey measures on store cleanliness, friendliness of personnel, loyalty cards and special deals. However, as these items are not linked in any way to NB entry or presence, we did not include them in our analysis. As for ‘special deals’, a reliability analysis revealed that this item could not be reliably merged with the price measure (Cronbach alpha = .621), which is not surprising, given that such deals at Lidl often involve non-branded products and ‘one-time’ offers outside of the regular assortment.

5. Not all panel members who received a request to participate in the survey in a specific wave actually responded: panelists were given the possibility to ‘opt-out’ after scanning their purchases. The typical response rate was 70%.

6. Extra analyses, in which the value and price scores are treated as separate dependent variables, reveal a similar impact of the NB introductions on both items.

7. Assortment image is thus measured using a single-item scale. As indicated by Diamantopoulos, Sarstedt, Fuchs, Wilczynski and Kaiser (2012), the use of single-item scales is deemed acceptable for more concrete constructs. Broniarczyk et al. (1998) also adopt a single-item measure of assortment perceptions.
EVOLUTION OF IMAGE SCORES FOR LIDL ACROSS SURVEY WAVES

Fig. 2. Evolution of image scores for Lidl across survey waves. *Y-axis represents mean item scores.

3.2.2. Operationalizations: explanatory variables

3.2.2.1. Store-level price and assortment variables. We use data from the complete national panel to calculate the weekly overall price and assortment for Lidl and the main competitors (i.e. Aldi, and the top-three mainstream retailers, which – together – account for more than 90% of total grocery spending). The observations are recorded at the level of the household, noting the chain where the shopping trip took place, and which (and how many) SKUs were purchased. For each SKU, we know its brand name and price, as well as the product category it belongs to (out of a set of 59 product categories defined by GfK). Assortment size of a store in a given week is then operationalized as the total number of different SKUs available at that store. Similar to Gielens (2012), to ensure the stability of this measure, we use a ‘running’ monthly horizon to ascertain whether an SKU is present (i.e. purchased by at least one household in the entire national panel in the course of the month up until and including that week). For Lidl, the average assortment size is 1770 SKUs (with a standard deviation (SD) of 340 across observation weeks), compared to 1440 SKUs (SD: 154) for Aldi, and to 6308 SKUs (SD: 305) for the mainstream stores – figures comparable to those reported in previous literature (see e.g. Nielsen, 2007; van Heerde et al., 2008). The chains’ actual overall price level at time t is computed as a weighted average of all SKU unit prices in that period. To allow for meaningful comparison across categories (where different volume units may apply: e.g. liters for soft drinks, kilograms for cereals), we first transform SKU unit prices within a category into indices. We do so by dividing the SKU’s observed price per volume unit (e.g. liter, kilogram), by the average unit price in the category across all weeks and stores (which comes down to a simple rescaling). For each store and week, we then aggregate these across SKUs in the category, and weigh the category-level price indices by the category’s average (market-wide) share-of-wallet to obtain an overall price index of the main stores (see van Heerde et al., 2008 for a similar procedure). The average price index for Lidl thus equals .719 (SD .066), a figure comparable to Aldi’s (average: .672, SD: .048), and substantially below the price indices of the main stores (average: 1.21, SD: .083).

3.2.2.2. HD proneness and (logged) image scores. The households’ inclination towards the HD format, which enters the image equations as a control (HDProne), is measured by the discount trip share, i.e. number of household trips to discounters (Aldi and Lidl in our data set) divided by the total number of household trips. The image equations further include lagged values as explanatory variables, to capture the ‘stickiness’ in consumer perceptions, and allow for carryover effects. Unfortunately, given the way the data were collected, the same household does not necessarily rate Lidl in every survey wave, and many households do not even provide ratings for (immediately) subsequent waves. Including only observations for which a same-household previous-wave score is available, would thus be very ‘costly’ in terms of data loss. Hence, in cases where no household-specific score is available, we use the aggregate rating across households in the previous wave for the lagged value and assortment image variables.8 A somewhat similar problem occurs in the SOW Eq. (2c), which includes value and assortment images as explanatory variables. Even though we only estimate that equation on households that participated in the survey, our survey data are bi-annual (not weekly) and, moreover, we do not have ratings for all households in each survey week. To solve this issue, we use an approach similar to van Heerde et al. (2008): we ‘assign’ weeks to the closest week in which a survey took place and, if no rating is available for the considered household in that survey week, use an imputed value based on the average rating in that week.

3.2.2.3. National brand entries and exits. To operationalize the variables related with NB introductions, we first classify the brands sold at Lidl into national brands and private labels. Given the large number of differently-named private labels at that chain, this is not an obvious task. We identify NBs as brands that (i) are not flagged (by GfK) as PLs, and (ii) are sold at each of the three traditional supermarket chains.9 Next, we record which of the NBs are available at Lidl, in which categories, and since when. We identify the introduction (exit) week of each brand, as the week in which an SKU purchase of that brand was first (last) recorded at Lidl in the entire national panel. Similar to Deleevensnyder et al. (2007), we obtain the between-store price gap for a NB introduced at the HD (PGap_between) as the ratio of its unit price (e.g. price per kg, price per liter) at the (top-three) traditional chains, relative to its unit price at Lidl. So, larger values of this measure indicate that Lidl charges less for the NB compared to its traditional supermarket competitors. To obtain the within-store price gap for a newly listed NB at Lidl (PGap_within), we divide its unit price at Lidl by the unit price of its PL counterpart at Lidl. Specifically, we compare the NB’s most popular SKU in a given subcategory to the most popular PL-SKU in that same ‘subcategory’, where subcategory is the most refined classification made available by GfK.10 As such, a higher level of PGap_within indicates that the NB is more expensive than comparable PLs at the HD. We follow Deleevensnyder et al. (2007) by using the logarithm of these (skewed) price ratios as explanatory variables in our models. The depth of the introduced NB’s line (LDepth) is operationalized as the number of SKUs at the time of introduction. We identify a NB as a ‘leading’ NB (Leader) if it is the top-selling brand in its category.

8 As an alternative, we considered a model with aggregate lagged scores throughout, even in cases where a household-specific lagged rating was available. The main pattern of results remained the same.

9 We further refine our procedure by discarding brands with very few purchase records at Lidl (i.e. less than ten observations throughout the observation period, or less than one observation per month between the first and the last purchase record), to avoid that coding errors are mistaken for NB listings. Likewise, given that they might have been available already by the end of the previous year (which we do not observe), brands identified before January 31, 2005 are not considered as new introductions. Finally, we do not use brands in fresh product categories (e.g. fruit and vegetables) or product categories outside of a typical supermarket assortment (e.g. electronic appliances).

10 If a NB was introduced at Lidl within several subcategories simultaneously, we calculated the gap in each subcategory, and then aggregated across these subcategories to the NB-category level.
Table 3
Overview of national brand entries.

<table>
<thead>
<tr>
<th>Product category (size)</th>
<th>NB intros</th>
<th>NB exits</th>
<th>NB leaders</th>
<th>Week intro</th>
<th>Wave intro</th>
<th>Examples of introduced national brands (NBs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dish detergent (13)</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>52–223</td>
<td>3–9</td>
<td>Finish, Sun, Per</td>
</tr>
<tr>
<td>Alcoholic drinks (81)</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>158–285</td>
<td>7–12</td>
<td>Corsendonk, Chimay, Maes Pils</td>
</tr>
<tr>
<td>Sandwich spread (24)</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>9–296</td>
<td>1–12</td>
<td>Callebaut, Du Vrai Sirop De Liège</td>
</tr>
<tr>
<td>Chocolate products (57)</td>
<td>12</td>
<td>6</td>
<td>0</td>
<td>4–255</td>
<td>1–10</td>
<td>Raffaello, Maltesers, Mon Chéri</td>
</tr>
<tr>
<td>Deodorants (10)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>11–177</td>
<td>5–8</td>
<td>Axe, Rexona</td>
</tr>
<tr>
<td>Skin care (37)</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>55–243</td>
<td>3–10</td>
<td>Dove, Clearasil, Sunlight</td>
</tr>
<tr>
<td>Pesticides (5)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>123–184</td>
<td>6–8</td>
<td>Baygon, Vapona</td>
</tr>
<tr>
<td>Oils and fats (17)</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>36–183</td>
<td>2–8</td>
<td>Bertolli, Becel, Planta</td>
</tr>
<tr>
<td>Salted snacks (28)</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>55–132</td>
<td>3–6</td>
<td>Duyvis, Pringles</td>
</tr>
<tr>
<td>Cheese (68)</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>10–270</td>
<td>1–11</td>
<td>Babybel, Chavroux, Chaumes, Leerdammer, Kiri</td>
</tr>
<tr>
<td>Dental care (9)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>59–155</td>
<td>3–7</td>
<td>Aquafresh, Signal</td>
</tr>
<tr>
<td>Non-alcoholic drinks (74)</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>22–290</td>
<td>2–12</td>
<td>7 Up, Coca-Cola, Lipton, Gini, Orangina, Oasis, Schweppes, Pepsi</td>
</tr>
<tr>
<td>Air fresheners (13)</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>58–132</td>
<td>3–6</td>
<td>Brise, Bref, Air Wick, Febrézé</td>
</tr>
<tr>
<td>Diapers &amp; feminine hygiene (20)</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>59–197</td>
<td>3–8</td>
<td>Always, Aldays, Pampers, OB, Tampax</td>
</tr>
<tr>
<td>All-purpose cleaners (23)</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>47–202</td>
<td>3–9</td>
<td>Carolin, Cillit Bang, Clif, Instanet, Mr Propre</td>
</tr>
<tr>
<td>Sweats (44)</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>9–304</td>
<td>1–12</td>
<td>Freudent, Cocomas, Look-O-Look, Stimorol, Tic Tac</td>
</tr>
<tr>
<td>Personal care (8)</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>58–122</td>
<td>3–6</td>
<td>Gillette, Oral B</td>
</tr>
<tr>
<td>Fine meat (87)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>23–258</td>
<td>2–11</td>
<td>Herta, Marcassou</td>
</tr>
<tr>
<td>Hot drinks (31)</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5–213</td>
<td>1–9</td>
<td>Nescafé, Ricoré</td>
</tr>
<tr>
<td>Laundry detergents (21)</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>10–176</td>
<td>1–8</td>
<td>Ariel, Disan, Lenor, Silan, Persil, Vanish</td>
</tr>
<tr>
<td>Other categories</td>
<td>11</td>
<td>4</td>
<td>2</td>
<td>10–267</td>
<td>1–11</td>
<td>Maggi, Knorr, Ola, Zwan, Granforno</td>
</tr>
</tbody>
</table>

a Only product categories with more than one NB introduction are displayed in a separate row. Size refers to the number of SKUs in Lidl’s assortment (average throughout the observation period).
   b Indicates the first and last week in which a NB intro was observed in the category.
   c Indicates the first and last survey wave in which a NB intro was observed in the category.

(i.e. one of the 59 GfK categories), market-wide. Prevalence of national brands in the hard-discounter’s assortment at the time of introduction (CUM_NBShare) is measured by the number of NB SKUs relative to the total number of SKUs at Lidl. Finally, a category’s purchase frequency (Cat_Freq) is derived from the entire national panel and indicates, among category buyers (i.e. households with at least one purchase in the category), the average number of times households engage in a purchase from the category per year.

Table 3 provides an overview of the NB introductions. 125 NB entries at Lidl are identified in the course of our observation period. As the table shows, these introductions are quite spread in time, with many of them occurring in the middle of the data period — leaving us with ample observations both prior to and after introductions. NB exits are far fewer in number (35), and mostly occur towards the end of the observation period, making it harder to reliably assess their (separate) effects. Hence, we include these exits as controls, and do not split up their effects depending on brand or category characteristics. The NB entries occur in a wide range of categories with varying purchase frequencies, from food products (e.g. chocolates, cheese, and alcoholic drinks) to many non-food items (e.g. laundry detergents, cleaning products, and diapers).

Summary statistics of the NB characteristics are given in Table 4. About 10% of the introduced brands are category leaders and represent the ‘big name’ in the category (e.g. “Coca Cola” in soft drinks, “Pampers” for diapers). The table further reveals quite some variation in brand-line depth, as well as in the price differential relative to competing mainstream stores and the private label offer. In all, the NB introductions represent a broad mixture of the moderator characteristics, with little overlap between them (as can be seen from the correlations in Table 4). Hence, the number and timing of NB entries, and the variation therein, provides an interesting basis for our study of image and SOW effects.

3.2.2.4. Endogeneity issues. To ensure that our estimates are unbiased, several possible endogeneity issues need to be considered. First, are the actual prices charged at Lidl exogenous to the value-scores and to the store’s share-of-wallet? Similar to van Heerde et al. (2008), we argue that store prices are highly unlikely to be adjusted based on individual households’ image perceptions – the dependent variables in our image equations – thereby precluding endogeneity.

Still, as a more formal check, we conduct a Hausman endogeneity test for price in the value image equation.11 As expected, no evidence of endogeneity is found. As for SOW, the week of introduction of the data virtually rules out store-level price adjustments within the same period. This is again confirmed by the Hausman test – pointing to price exogeneity in the SOW equation.

A second question concerns the possible endogeneity of the NB entry and exit decisions. Surely, NB introductions or delistings may be affected by consumers’ perceptions of, and SOW spent at, the HD store – the key question being whether such reverse causal effects occur within the same data period. For the SOW equation, estimated on weekly data, this will not be the case. To ascertain exogeneity for the bi-annual image equations, we again conduct a Hausman endogeneity test. Given the number of entries and exits, we do not carry out this test for each NB introduction or removal decision separately but, rather, consider the number of introductions and deletions at a given point in time – an approach similar to Gielens (2012).12 For both the value and the assortment perceptions, we find that endogeneity is not an issue.

4. Empirical results

4.1. Impact of NB introductions on value and assortment image

Table 5A summarizes the estimation results. The correlation between the two error terms equals .433 – pointing to a halo effect. Turning to the parameter estimates, we find the expected positive coefficient for HD proneness – capturing the fact that frequent visitors of hard-discounter stores hold more favorable perceptions of Lidl’s value (α5 = .300; p < .01) and assortment (β5 = .870; p < .01) to start with. The positive coefficient of lagged assortment ratings (β5 = .299; p < .01) reflects previous findings that images are

---

11 We use the previous-wave Lidl price, along with lagged NB entries and exits, and a time trend, as instruments. The use of lagged values as instruments is quite common (e.g. Dhar & Hoch, 1997; Villas-Boas & Winer, 1999) and acceptable here, given the absence of autocorrelation. Note that since price does not enter the assortment-image equation, no endogeneity test is needed for that equation.

12 For both variables, we use as instruments their lagged values in the previous wave, as well as the lagged value- and assortment perceptions, and a time trend. Adding the residuals to the image equations shows that they do not have a significant effect.
### Table 4

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category purchase frequency</td>
<td>12.530</td>
<td>5.790</td>
<td>1</td>
</tr>
<tr>
<td>NB share at Lidl</td>
<td>0.318</td>
<td>0.032</td>
<td>0.020</td>
</tr>
<tr>
<td>Leader brand</td>
<td>0.113</td>
<td>0.318</td>
<td>0.172</td>
</tr>
<tr>
<td>NB price gap between Lidl and mainstream stores</td>
<td>1.166</td>
<td>0.261</td>
<td>0.324</td>
</tr>
<tr>
<td>Price gap within Lidl</td>
<td>2.691</td>
<td>1.847</td>
<td>0.124</td>
</tr>
<tr>
<td>NB Line depth</td>
<td>1.858</td>
<td>1.177</td>
<td>0.085</td>
</tr>
</tbody>
</table>

### Table 5

**Panel A: Image Equations**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expected effect</th>
<th>Parameter</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intro NB</td>
<td>(+)</td>
<td>( \mu_i )</td>
<td>0.014</td>
<td>0.011</td>
<td>0.228</td>
</tr>
<tr>
<td>Intro NB × purchase frequency</td>
<td>(+)</td>
<td>( \mu_i )</td>
<td>0.004</td>
<td>0.012</td>
<td>0.763</td>
</tr>
<tr>
<td>Intro NB × leader brand</td>
<td>(+)</td>
<td>( \mu_i )</td>
<td>0.311***</td>
<td>0.096</td>
<td>0.001</td>
</tr>
<tr>
<td>Intro NB × price gap Between</td>
<td>(+)</td>
<td>( \mu_i )</td>
<td>0.092</td>
<td>0.226</td>
<td>0.684</td>
</tr>
<tr>
<td>Intro NB × price gap within</td>
<td>(+/-)</td>
<td>( \mu_i )</td>
<td>0.180**</td>
<td>0.074</td>
<td>0.015</td>
</tr>
<tr>
<td>Intro NB × Cum NB SKU Share</td>
<td>(+/-)</td>
<td>( \mu_i )</td>
<td>0.083***</td>
<td>0.019</td>
<td>0.001</td>
</tr>
<tr>
<td>Exit NB</td>
<td></td>
<td>( \alpha_1 )</td>
<td>0.013</td>
<td>0.020</td>
<td>0.512</td>
</tr>
<tr>
<td>1-period lagged value image</td>
<td></td>
<td>( \alpha_5 )</td>
<td>0.099</td>
<td>0.101</td>
<td>0.330</td>
</tr>
<tr>
<td>HD proneness</td>
<td></td>
<td>( \alpha_5 )</td>
<td>0.300*</td>
<td>0.160</td>
<td>0.061</td>
</tr>
<tr>
<td>Overall price index Lidl</td>
<td></td>
<td>( \alpha_5 )</td>
<td>3.110</td>
<td>3.928</td>
<td>0.429</td>
</tr>
<tr>
<td>Assortment image</td>
<td></td>
<td>( \lambda_1 )</td>
<td>0.082***</td>
<td>0.029</td>
<td>0.004</td>
</tr>
<tr>
<td>Intro NB</td>
<td>(+)</td>
<td>( \lambda_1 )</td>
<td>0.005</td>
<td>0.007</td>
<td>0.410</td>
</tr>
<tr>
<td>Intro NB × purchase frequency</td>
<td>(+)</td>
<td>( \lambda_4 )</td>
<td>0.272</td>
<td>0.243</td>
<td>0.263</td>
</tr>
<tr>
<td>Intro NB × leader brand</td>
<td>(+)</td>
<td>( \lambda_5 )</td>
<td>0.129</td>
<td>0.055</td>
<td>0.020</td>
</tr>
<tr>
<td>Intro NB × Cum NB SKU share</td>
<td>(+/-)</td>
<td>( \lambda_6 )</td>
<td>0.007</td>
<td>0.025</td>
<td>0.781</td>
</tr>
<tr>
<td>Exit NB</td>
<td></td>
<td>( \lambda_7 )</td>
<td>0.006**</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td>1-period lagged assortment image</td>
<td></td>
<td>( \lambda_7 )</td>
<td>0.299***</td>
<td>0.070</td>
<td>0.000</td>
</tr>
<tr>
<td>HD proneness</td>
<td></td>
<td>( \lambda_7 )</td>
<td>0.870***</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall assortment Lidl (1000 SKUs)</td>
<td></td>
<td>( \lambda_7 )</td>
<td>1.400</td>
<td>0.800</td>
<td>0.093</td>
</tr>
</tbody>
</table>

**Panel B: share-of-wallet (SDW) equation**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Predicted effect</th>
<th>Parameter</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assortment image</td>
<td>(+)</td>
<td>( \gamma_4 )</td>
<td>0.00679**</td>
<td>0.00339</td>
<td>0.045</td>
</tr>
<tr>
<td>Value image</td>
<td>(+)</td>
<td>( \gamma_5 )</td>
<td>0.01290**</td>
<td>0.00451</td>
<td>0.004</td>
</tr>
<tr>
<td>Intro NB</td>
<td>(+/-)</td>
<td>( \gamma_5 )</td>
<td>-0.00092**</td>
<td>0.00005</td>
<td>0.404</td>
</tr>
<tr>
<td>Intro NB × purchase frequency</td>
<td>(+)</td>
<td>( \gamma_6 )</td>
<td>0.00030</td>
<td>0.00096</td>
<td>0.756</td>
</tr>
<tr>
<td>Intro NB × leader brand</td>
<td>(+)</td>
<td>( \gamma_7 )</td>
<td>0.00140</td>
<td>0.00163</td>
<td>0.388</td>
</tr>
<tr>
<td>Intro NB × price gap between</td>
<td>(+/-)</td>
<td>( \gamma_8 )</td>
<td>-0.00092*</td>
<td>0.00049</td>
<td>0.061</td>
</tr>
<tr>
<td>Intro NB × price gap within</td>
<td>(+/-)</td>
<td>( \gamma_9 )</td>
<td>0.00018</td>
<td>0.00025</td>
<td>0.460</td>
</tr>
<tr>
<td>Intro NB × Cum NB SKU share</td>
<td>(+/-)</td>
<td>( \gamma_9 )</td>
<td>0.00144***</td>
<td>0.00045</td>
<td>0.001</td>
</tr>
<tr>
<td>Exit NB</td>
<td></td>
<td>( \gamma_9 )</td>
<td>0.00016</td>
<td>0.00018</td>
<td>0.364</td>
</tr>
<tr>
<td>1-period lagged SDW</td>
<td></td>
<td>( \gamma_9 )</td>
<td>0.37156***</td>
<td>0.01086</td>
<td>0.000</td>
</tr>
<tr>
<td>Nr. SKUs Lidl (1000 SKUs)</td>
<td></td>
<td>( \gamma_9 )</td>
<td>0.07000***</td>
<td>0.01350</td>
<td>0.000</td>
</tr>
<tr>
<td>Nr. SKUs Aldi (1000 SKUs)</td>
<td></td>
<td>( \gamma_9 )</td>
<td>0.00200</td>
<td>0.01550</td>
<td>0.878</td>
</tr>
<tr>
<td>Nr. SKUs mainstream (1000 SKUs)</td>
<td></td>
<td>( \gamma_9 )</td>
<td>-0.00800**</td>
<td>0.00337</td>
<td>0.025</td>
</tr>
<tr>
<td>Price Lidl</td>
<td></td>
<td>( \gamma_9 )</td>
<td>-0.06751***</td>
<td>0.02387</td>
<td>0.005</td>
</tr>
<tr>
<td>Price Aldi</td>
<td></td>
<td>( \gamma_9 )</td>
<td>-0.02289</td>
<td>0.02186</td>
<td>0.295</td>
</tr>
<tr>
<td>Price mainstream stores</td>
<td></td>
<td>( \gamma_9 )</td>
<td>-0.02860**</td>
<td>0.01268</td>
<td>0.023</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>( \gamma_9 )</td>
<td>0.02471</td>
<td>0.03670</td>
<td>0.501</td>
</tr>
<tr>
<td>Riquare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# observations = 243708 (households by shopping weeks)

Estimated in STATA with a seemingly unrelated regression for unbalanced panel with random effects.

# observations = 3113 (households by survey waves in which they participated).

\( * p < .01, ** p < .05, * * p < .1 \).

\( * * * p < .01, * * * p < .05, * * p < .1 \).

\( a \) Moderator variables mean-centered prior to calculating interaction variables. **\( p < .01, * * p < .05, * p < .1 \).
While NB delistings do not alter perceived ‘value-for-money’, they have a significant negative effect on the HD’s assortment image ($\beta_2 = -0.36; p < 0.01$).

Focusing on the NB introductions, we find the main effect to be positive and significant in the assortment equation ($\lambda_1 = 0.082; p < 0.01$), indicating that ‘on average’, NB listings lead to more favorable assortment perceptions. We also obtain a positive interaction effect with brand line depth ($\lambda_3 = 0.129; p < 0.05$), suggesting that introductions involving more SKUs further enhance the HD’s perceived assortment variety. In the value image equation, although the main effect of introductions is insignificant ($\mu_1 = 0.014; p > 1.0$), Table 5A reveals that the characteristics of introductions may significantly alter the magnitude and direction of their effect. Additions of leading NBs ($\mu_4 = 0.311; p < 0.01$) more positively contribute to the HD’s value image. As for price, we find that within — rather than across-stores comparisons play a role: while the between-stores price gap shows no effect ($p > 10$), the NB-PL price differential within Lidl exerts a strong negative moderating effect ($\mu_6 = -180; p < 0.05$). Interestingly, the cumulative presence of NBs at the discounter also matters, and negatively moderates the impact of further NB additions on store value perceptions ($\mu_8 = -0.083; p < 0.01$).

In all, it seems that NB listings may exert an important influence on the HD’s image, and differently so depending on their specific characteristics. However, Table 5A provides only a cursory assessment: to fully appreciate the interaction effects of NB characteristics, we need to combine them with other moderators and their coefficients (Jaccard & Turrisi, 2003) — something we’ll turn to in the implications section below.

4.2. Impact on SOW

Table 5B presents the results for the SOW equation. The store’s actual marketing mix has the expected effect on share-of-wallet: Lidl’s overall price level has a significant negative impact ($\gamma_8 = -0.068; p < 0.01$), whereas the effect of its assortment size is positive ($\gamma_9 = 0.070; p < 0.01$). As anticipated, larger assortments at mainstream stores hurt Lidl’s purchase share ($\gamma_{11} = -0.008; p < 0.05$). The same goes for price increases at those stores ($\gamma_6 = -0.028; p < 0.05$), possibly because, with inelastic demand, they enhance the Euro-value of baskets purchased at those stores, thus lowering Lidl’s share of spending. Weekly price and assortment levels of Aldi do not have a significant impact, in line with the observation that households mainly cross-shop at one HD and mainstream stores, rather than among HDs (Gijsebcrechts, Campo & Nisol, 2008). The coefficient of NB exits does not reach significance ($p > 10$).

As expected, the store’s assortment image ($\gamma_4 = 0.007; p < 0.05$) and value image ($\gamma_5 = 0.013; p < 0.01$) exert a significant positive effect on SOW and, as such, mediate the impact of NB listings on consumers’ spending share at the HD. In addition, we do find some significant direct effects of NB introductions on SOW: a negative main NB introduction effect ($\theta_1 = -0.00092; p < 0.05$), and a significant effect for two moderating characteristics: the purchase frequency of the category where the NB is introduced ($\theta_2 = 0.00010; p < 0.05$), and the presence of NBs in the HD’s assortment at the time of introduction ($\theta_3 = 0.00144; p < 0.01$). However, as we show below, the magnitude of these direct effects remains modest. Hence, as anticipated, NB introductions influence households’ spending share at the store primarily through their impact on store perceptions.13 Zooming in on these mediators, we find that a change in the ‘value-for-money’ rating is almost twice as influential on SOW as a same change in the assortment rating — in line with the premise that HD shoppers are mainly looking for good deals.

4.3. Implications

The results above suggest that (i) adding NBs on the discounter’s shelves may lead to a more favorable assortment and value image but only under well-chosen conditions, (ii) the direction of the effect for a given NB entry may well be opposite for the two image components, and (iii) the ultimate impact of NB introductions on the store’s SOW will be the net outcome of both direct effects and indirect ones, through perceptual changes.

To get a better grip on the impact and significance of NB entries, which is spread across both main and interaction coefficients in both the image equations and the SOW equation, we proceed as follows. We first compile alternative entry ‘events’ as combinations of key moderator characteristics, i.e. category leader or not, few vs. many introduced SKUs, and small vs. large price gap relative to Lidl PLs (keeping other moderators at their mean level). For the within-store price gap, we consider ‘low’ and ‘high’ values close to the lower and upper quartile in the data. For number of SKUs, we set the high value at 5 SKUs – a level still well within the data range – to more clearly show its effect. Next, for each such event, we use our model parameters to calculate the resulting change in value-image, assortment-image and store-SOW. Because the model parameters are estimated with uncertainty, we take 1000 random draws from their multivariate sampling distributions (characterized by the mean and variance-covariance matrix of the estimates in the image and SOW models), and do the calculations for each draw. We then use these draws to assess the mean introduction effect and both lower and upper 2.5 percentiles.

Fig. 3 summarizes the results. For each introduction case, the associated perception shifts are depicted in the two left panels. The right panel shows the corresponding changes in SOW. To save space and facilitate comparison, we show the impact for leading vs. non-leading NBs in the same plot.

An interesting pattern emerges. First, the figure confirms that NB introductions can significantly alter the HD’s value and assortment image, and its ensuing share-of-wallet. Some NB introductions generate a positive effect across-the-board. Specifically, NB additions at a small price premium over the prevailing PLs, and with many SKUs (Fig. 3A), significantly enhance consumer perceptions of the hard-discounter’s value and assortment — irrespective of whether the brand is a category leader or not. These image improvements, in turn, translate into a significantly higher share-of-wallet, especially for brands that are category leaders (+ .78% share points, compared to + .41% for non-leaders). Such positive effects, however, only materialize for specific, favorable combinations. For instance, narrowing the line of introduced SKUs (Fig. 3B) takes away the positive assortment-image effect and, thereby, the increase in SOW. Likewise, again starting from the favorable scenario in Fig. 3A but increasing the within-store price gap, reduces (removes) the value-perception improvement for leading (non-leading) brands, and takes away the positive SOW effect for non-leaders (Fig. 3C). So, for NB listings to exert a positive impact on both image dimensions or SOW, they must be selected and managed carefully.

Moreover, not only may NB additions fail to generate the desired image and SOW effects, they even constitute an imminent threat. This can be seen from the scenario in Fig. 3D, where, unless it is a category leader, the introduction of a narrow-line NB with a sizeable price gap, entails a significant reduction in all three outcome metrics. So, not only do such entries hamper the HD’s value image, they further reduce its assortment perception, resulting in a significant drop in SOW.

From Fig. 3, it appears that the changes in SOW closely reflect the underlying pattern of changes in value and assortment image. In fact, across the cases in Fig. 3 in which at least one of the two effects is significant, the ratio of the (absolute value of the) indirect effect to the sum of the (absolute values of the) direct and indirect effects, amounts to 85% on average. Hence, though NB entries do exert a

---

13 Over and above their effect on store price and assortment size (captured in the overall price and assortment variables).
direct influence on SOW, the dominant part of the change comes from the indirect effect of value and assortment perceptions. This is especially true for the introduction of category leaders, where well over 90% of the spending-share change materializes through store images. This confirms that mindset metrics, indeed, constitute important constructs through which NB listing affects store performance.

The model estimates also revealed an effect of the cumulative presence of NBs at the HD. To further pursue this, we re-simulate the impact of the introduction events for higher NB assortment shares (i.e. 38%, the mean level plus two standard deviations). We find that for NB listings that enhanced the HDs value perceptions at the current NB-presence (i.e. 30% of the overall HD assortment; effect shown in Fig. 3A and B), the impact is now smaller, but the relative drop is limited (i.e. a less than 2% reduction, from .363 to .357). For introductions that were harmful in the first place (Fig. 3C and D), the negative value-image effect is exacerbated at higher NB shares (i.e. value-ratings further drop by 9%). Interestingly, when it comes to spending share, these value-image changes are compensated by the (positive) direct effect of cumulative brand presence (see Table 5B), thus leaving the SOW effect of the introductions unchanged. So, whereas higher NB presence reinforces the negative value-image consequences of yet more NB introductions, our data show no such effects for spending.

![Fig. 3. Impact of nb introductions](image-url)
4.4. Robustness checks

To ensure the validity of our findings, we conduct several robustness checks. First, we estimate the image models allowing NB introductions to produce trend shifts in perceptions, over and above the level shifts captured by the step dummies. We find the trend coefficients to be insignificant (which is not too surprising, given that long-term effects are already taken up by the lagged perception coefficients). We also add pulse variables for brand entries and exits to test for any temporary shifts at the time of NB introduction/delisting, capturing the possibility that consumers revise their images of the store at first, but then partly revert to their previous perceptions. Again, we do not find evidence for such effects.

Next, to rule out that our SOW parameters are an artefact of the value and assortment imputation, we re-estimate the SOW equation using only household-weeks in which an actual survey rating was available. Though these are, by construction, weeks in which the household visited a Lidl store at least once (and, therefore, do not constitute a representative subset, something we return to in the discussion section), such estimation is still instructive to assess the image effects on size of the baskets bought at Lidl. We find the results to be similar, with a strong positive effect for assortment image and an even stronger effect for value image.

As a further refinement, we examine different specifications for the impact of brand line depth. For one, we include its square as an extra moderator, to accommodate the possibility of an inverse-U effect (overly deep lines possibly creating confusion and overload in shoppers’ minds). We find no evidence for such an effect — the added variable being insignificant. As an alternative operationalization, we include the number of introduced SKUs relative to the total number of SKUs in the HD’s category offer. Unlike its absolute counterpart, this relative measure does not show a significant effect — suggesting that, indeed, it is the absolute presence on the HD shelves that drives shoppers’ perceptions.

We further check the robustness of the results by including different measures for the stores’ marketing mix in the image and SOW equations. Specifically, we replace the ‘total number of SKUs in the store’ by ‘total number of brands’, and use ‘prices relative to competing stores’ instead of the absolute price measures. The key results (i.e. impact of NB entries on value and assortment image, and image effects on SOW) remain unchanged. We also add the share of PLs in the main stores’ assortment as an extra control in Lidl’s SOW equation. The effect of this variable is insignificant, whereas the impact of NB entries and store images remains the same as before.

Finally, because households’ weekly share-of-wallet at Lidl is rather volatile, we re-run the SOW model after aggregating the data to the four-week level. While this makes the lagged SOW effect much more important, and the own- and competitive marketing mix less significant, it does not change the pattern of results for our purposes: the assortment and value perceptions remain highly significant (with the effect of the latter twice as high), and the direct effect of NB introductions on SOW stays mostly insignificant. In all, these checks strengthen our confidence in the findings.

5. Discussion, implications and future research

5.1. Discussion

Being positioned as lean, cheap, and strongly private-label focused originally, hard-discounters have started to accept more NBs in their assortment, and are expected to continue going down that road (Planet Retail, 2010). In so doing, they count on getting even better inroads into the mainstream supermarkets’ customer segments. Previous research by Deleersnyder et al. (2007) revealed that national brands may, indeed, enhance HD sales in the categories where they are introduced. However, being a departure from the HD’s original business model, NB listings may have an effect well beyond the category in which they occur — influencing the store as a whole.

Our study offers empirical evidence for the impact of NB introductions on the HD’s overall value and assortment image, and its ensuing share of the customers’ wallet. We find that, while potentially rewarding, such NB introductions do not come without risk. Whereas carefully managed NB listings may produce the hoped-for ‘double whammy’ effect on reputation and spending, ill-selected NB introductions are bound to severely hurt the HD’s aura of ‘good value-for-money’ — cutting into its core competitive advantage and leading to a notable decline in SOW.

5.1.1. NB listings and store image

We find that leading national brands typically enhance the HD’s ‘value-for-money’ image. Hence, even if category leadership is not a driving force when it comes to increasing category sales (Deleersnyder et al., 2007), leading brands have signaling power at the store level. This is consistent with the common wisdom that “a retailer with a relatively low image might be able to improve this image by associating it with a more favorably evaluated brand or manufacturer image” (Jacob & Mazursky, 1984, p.121). Non-leading NBs, more often than not, do not generate such positive effects, and may even hurt the HD’s value positioning. Like many Western-European countries, the Belgian market is a “PL-mature” country, in which consumers attribute smaller quality gaps to NBs (Steenkamp et al., 2010) — especially to less-innovative, less-advertised NBs, which would typically not be category leaders. Recent studies suggest that such small perceived quality differences also apply to the hard-discounter’s private label (Testaankoop, 2012). The perceived quality superiority of those NBs may simply be too small to induce a meaningful shift in the HD’s quality image. Combined with the fact that their price remains higher than that of the HD’s ‘competing’ PL product, this explains the negative effect of non-leading NBs on perceived value.

Interestingly, category leadership is not a strength when it comes to assortment perceptions. Rather, in that case the presence of a line of SKUs is important. While NB entries with multiple SKUs enhance the HD’s assortment perception, we find that single-SKU entries may even produce a negative effect. A plausible explanation is that such NB entries create a ‘contrast’ with mainstream stores, where the brand’s line is more extensive. This may reinforce the perception, especially among users of the NB, that HD assortments are indeed quite narrow and offer little choice variety, thereby further lowering the HD’s assortment image.

Like Deleersnyder et al. (2007), we find price to play an important role, albeit for different reasons. First, the nature of the effect on our outcome metrics is different. HD’s value image appears to be shaped by within-store rather than between-store price gaps: while the HD price difference between the HD and its mainstream competitors does not exert a significant impact, the price distance relative to the discounter’s private label offer is quite influential. This suggests that consumers predominantly rely on in-store cues (rather than memory-based processes) for HD value-image formation (Hamilton and Chernev, 2010a). Second, the direction of the effect is different. While Deleersnyder et al. (2007) observe a positive influence on category sales (possibly because of a switch to higher-priced NB items), we find that larger price gaps between the NB price and its PL counterpart threaten the HD’s overall value image. Having (more expensive) NBs on the HD’s shelves may take away the certainty that ‘whatever choice is made, it is going to be cheap,’ and thus alienate consumers who, while appreciating the quality of its PL offer, prefer an EDLP (HD) store as a substitute for price search (Bell & Lattin, 1998).

5.1.2. Mindset metrics and share-of-wallet

When it comes to SOW, we find that the impact of NB entries, for the larger part, comes as a result of changes in shoppers’ value and assortment perceptions. Not surprisingly, in this case the HD’s value
reputation carries the highest weight, suggesting that especially NB entries with a value-enhancing (value-reducing) effect will result in higher (lower) spending shares for the store. Particularly for introduction of category leaders, the change in SOW stems from an adjustment of the hard-discounter’s value image. This not only underscores the signaling role of these leading brands (Hamilton & Chernev, 2010a), it also confirms the importance of considering mindset metrics (Srinivasan et al., 2010). As indicated by Srinivasan et al. (2010), inclusion of such metrics may enhance our understanding of how marketing actions work and allow to discern effects that would otherwise go unnoticed. We find this premise to be true in our setting. Whereas both the impact of NB entries on store perceptions and the subsequent impact of store perceptions on SOW are significant and robust, the direct effects of NB introductions (and their moderators) on SOW are modest. Moreover, dropping the image components from the SOW model leaves these effects virtually unchanged: the NB introduction variables do not ‘pick up’ the effect of the omitted mindset metrics, possibly because of the complex dynamics, and countervailing influences on value and assortment positioning. Hence, considering the intermediate image effects is enlightening and provides guidance to HD retailers on NB selection, which would not be available otherwise.

5.1.3. Boundary conditions
Interestingly, even within the narrow window observed in our data, we find the impact of NB listings on the HD’s value perceptions to become less positive (or, for that matter, more negative) as more NBs are available on the HD shelves. Though the effect is small, and – at least for the observed NB presence – does not show up in the spending share for the store, it does call for caution and suggests that there is a limit to the number of NBs that can be successfully introduced by hard-discounters.

5.2. Managerial implications
Following recent press messages, hard-discounters seem to feel that NB introductions are ‘the way to go’, and increasingly adopt them as a tool to fight traditional retailers. Factual evidence appears to support this practice, to the extent that – indeed – NB presence enhances category sales (Deleersnyder et al., 2007). If, however, as suggested by Nielsen (2008), “it’s all in the mind,” then hard-discounters have an interest in revisiting the appropriateness of becoming more NB-oriented, and in considering the impact of NB introductions on their overall positioning.

On the one hand, we find that introduction of (especially leading) NBs may significantly contribute to a more favorable value perception of the HD store, provided they are offered at a reasonable price differential from the discounter’s PLs. What constitutes a ‘reasonable’ price differential, however, strongly depends on the brand position. In our application, leading brands can charge prices up to 2.5 times as high as those of the HD’s private labels before negative value-image effects set in, whereas prices of non-leading NBs should remain below 1.7 times the prevailing PL prices. At the same time, Deleersnyder et al. (2007) showed that a price premium over PLs may be instrumental in enhancing the HD’s category sales. They obtained such effects in a setting with NB prices twice as high as those of the PLs, on average. Taken together, this suggests that especially leading NBs hold the promise of success – allowing to maintain a large enough price gap to preserve the NB’s equity status and reap revenue benefits, without alienating HD shoppers in need of good prices across-the-board. For the NB entry to enhance the HD’s assortment perception, it must come with a sufficiently deep line, presenting a minimum variety-of-choice to shoppers interested in the NB. Given their limited shelf space, this is a requirement that HD retailers do not take lightly.

On the other hand, our study documents that ‘wrong’ NB introductions exert an unfavorable value image effect that is non-negligible. This is particularly true for NBs that are not category leaders — the large majority of actual introductions. For instance, given our estimated coefficients, introducing a shallow line of a non-leading NB with a NB–PL price gap towards the high end of what is observed in the data, may entail a .07 point decline in value image ratings (and a .05 point drop in assortment image ratings). This is an important effect in view of the observed distribution of these ratings, especially since the image change would linger over future periods. Moreover, following one such introduction, customers’ share-of-wallet may decline by about 0.23 percentage points in the short run (and almost .4 percentage points in the long run) — a non-negligible figure. Hence, HDs must be cautious not to alienate their current clientele and enhance immediate category sales to the detriment of their overall reputation and market position.

In all, these findings suggest that though potentially powerful, there are limits to what can be achieved with NB introductions — especially for HDs that already carry a large share of manufacturer brands in their assortment. Rather than further reducing their PL focus, HDs may opt for strategies more in line with their core positioning. For one, given the positive SOW effect of assortment size, HDs may explore ways to enhance the variety of their PL offer within the limits of available shelf space. They could reduce the number of facings per SKU, use more ‘efficient’ shelf displays or opt for higher SKU rotation — a practice successfully adopted by some retail firms (Mantrala, Levy, Kahn, Fox, Gaidarev, Dankworth, & Shah, 2009). Recently, to appeal to ‘upscale-oriented’ shoppers, hard-discounters have also ventured into diversification of their own PL lines, adding premium umbrella brands like “Delieieux” and “Deluxe” (Lidl), or “Excellence” (Aldi-Nord) in the Belgian market (Planet Retail, 2010). Uncovering the reputation and SOW effects of these strategies is an interesting future research topic.

5.3. Future research
Clearly, our study has limitations that lead the way for further research. First, in exploring the impact of NB introductions on household perceptions, we are limited by the nature of our data. It is possible, for instance, that the impact of NB line depth follows a nonlinear shape and that — even though the observed set of NB introductions is quite diverse — our data cover a range where this curve is flat. Similarly, despite the observed variation, the NB price differential between the HD and mainstream competitors may have remained too small to generate a noticeable perception effect. Hence, our findings may not apply to NB entries with more extreme characteristics than those considered here. Furthermore, though the Belgian market constitutes an interesting setting, it may be worthwhile to study the generalizability of our findings in other countries that differ in HD market position and NB offering.

Second, while we did have individual-level data, we typically had only few subsequent survey observations on store value and assortment image per household. This forced us to use less than optimal measures for the lagged perceptual scores in the image equations, such that the dynamics of these effects should be considered with caution. Also, our measures did not allow to distinguish the separate impact of NB introductions on the store’s quality and price perceptions — something we leave for future study. Moreover, the survey data were collected among panelists visiting the HD store at least occasionally. Hence, our findings only pertain to consumers with a non-zero propensity to patronize the HD format in the first place. Still, between 70 and 80% of the national sample of panelists report having visited a hard-discounter chain at least once in any given year. Moreover, with HD trip shares ranging from a low 1% to a high 100% (mean 53.1%, standard deviation 23.1 percentage points), our

---

14 Formally, a 1 percentage point change in SOW would yield a ‘long term’ effect of 1/(1-lagged coefficient) = 1.6 percentage points.
survey panelists appear representative of the population of shoppers. A remaining question, then, is whether the HD’s core customers are more likely to appreciate (or, rather, dislike) increased NB listings?

To further explore this question within the limitations of the data at hand, we re-ran the perception model including an interaction between the NB introduction variable, and the household’s general inclination to patronize a hard-discounter format (i.e. HD proneness). While the coefficient is negative in both image equations, it does not reach significance — suggesting that the HD’s core clientele does not perceive such additions as less appealing per se. Future studies using more (frequent) image ratings by household could try to better tease out reaction heterogeneity and to provide a more accurate picture of the long term effects.

Third, given the number of NB (de)listings at hand, we could not discern factors driving the impact of NB exits, and were limited in the number of moderators that we could consider for NB entries. As indicated by Steenkamp et al. (2010), the quality perception of NB versus PLs, and consumers’ willingness to pay for NBs, vary by category. Moreover, categories differ in their signaling value for store image (Lourenço et al., 2012). While the coefficient of category purchase frequency had the expected positive sign, it failed to reach statistical significance, which may be due to data limitations or because other category characteristics play a more dominant role. Future research may further tap into the type of categories where NB introductions enhance the overall value and assortment perception of the HD offer. In a similar vein, for lack of data, we could not assess the impact of any HD promotions of the listed NBs, nor look into the effects of brand characteristics such as advertising intensity, innovativeness, or package design (Deleersnyder et al., 2007; Gielens, 2009). Intra- and interformat competition among discounters and supermarkets. Marketing Science, 29(3), 456–474.


The influence of ad-evoked feelings on brand evaluations: Empirical generalizations from consumer responses to more than 1000 TV commercials

Michel Tuan Pham a,⁎, Maggie Geuens b,c, Patrick De Pelsmacker d,b

a Columbia University, Graduate School of Business, 3022 Broadway, New York, NY 10027, United States
b Ghent University, Tweekerkenstraat 2, 9000 Ghent, Belgium
c Vlerick Business School, Reep 1, 9000 Ghent, Belgium
d University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

A R T I C L E   I N F O

Article history:
First received in 25, January 2012
and was under review for 5 months
Available online 20 June 2013

Guest Editor: Marnik C. Dekimpe
Area Editor: Hans Baumgartner

Keywords:
Advertising
Emotions
Feelings
Involvement
Motivation
Empirical generalizations

A B S T R A C T

It has been observed that ad-evoked feelings exert a positive influence on brand attitudes. To investigate the empirical generalizability of this phenomenon, we analyzed the responses of 1576 consumers to 1070 TV commercials from more than 150 different product categories. The findings suggest five empirical generalizations.

1. Revisiting the effects of ad-evoked feelings on brand evaluations

Twenty-five years ago, an influential series of controlled lab studies seemed to indicate that feelings evoked by advertisements have a positive influence on consumers’ brand attitudes (Aaker, Stayman, & Hagerty, 1986; Batra & Ray, 1986; Edell & Burke, 1987; Holbrook & Batra, 1987). Everything else being equal, participants in these studies typically reported more favorable brand attitudes after viewing ads that elicited pleasant feelings than after viewing ads that elicited less pleasant feelings. These early findings have been replicated in a large number of studies with both television commercials and print advertisements (e.g., Baumgartner, Sujan, & Padgett, 1997; Burke & Edell, 1989; Derbaix, 1995; MacInnis & Park, 1991; Miniard, Bhatia, Lord, Dickson, & Unnava, 1991; Morris, Woo, Geason, & Kim, 2002). In a meta-analysis, Brown, Homer, and Inman (1998) found that the average correlation between ad-evoked feelings and brand evaluations was around \( r = .35 \) to \( .40 \), an effect that typically would be considered “medium to large” (Cohen, 1988).

The finding that brand attitudes can be substantially influenced by the mere emotional pleasantness of brand advertisements has obvious managerial implications: It would suggest that everything else being equal, marketers should generally try to improve the emotional pleasantness of their advertising. However, before this finding and its marketing implications can be accepted at face value and acted upon by managers, it is important to assess its empirical generalizability within the real world of brand advertising. To this end, two sets of issues need to be addressed.

First, because previous studies have focused primarily on the theoretical explanation of the phenomenon and therefore prioritized concerns for internal and construct validity, these studies have tended to compromise the external validity of their methodologies. For example, an important limitation of previous studies is that they were usually conducted among college students as opposed to more broadly representative samples of consumers (e.g., Machleit & Wilson, 1988; Madden, Allen, & Twible, 1988; Moore & Hutchinson, 1983; Stayman & Aaker, 1988). However, it has been observed that the correlation between ad-evoked feelings and brand attitudes tends to be stronger among students than among nonstudents (Brown & Stayman, 1992).

In addition, most of the classic studies involved rather limited pools of advertisements. For example, Edell and Burke (1987) examined ten

⁎ Corresponding author. Tel.: +1 212 854 3472; fax: +1 212 316 9214.
E-mail addresses: tdtp4@columbia.edu (M.T. Pham), Maggie.Geuens@UGent.be (M. Geuens), Patrick.Depelsmacker@ua.ac.be (P. De Pelsmacker).
and six commercials across studies. The limited pool of advertisements used in prior studies raises issues of potential selection bias. In particular, if the ad stimuli were specifically selected because of their emotional richness, prior studies may overstate the influence of ad-evoked feelings on brand evaluations in the real world. This potential selection bias is compounded by the fact that a number of previous studies relied on fictitious or unfamiliar ads and brands (e.g., Mackenzie & Lutz, 1989; Miniard, Bhatla, & Rose, 1990; Olney, Holbrook, & Batra, 1991; Park & Young, 1986), which tends to further amplify the observed influence of ad-evoked feelings on brand evaluations (e.g., Brown et al., 1998; Derbaix, 1995; see also Bakamitsos, 2006; Ottati & Isbell, 1996).

Finally, most of the previous studies (e.g., Batra & Ray, 1986; Burke & Edell, 1989; Edell & Burke, 1987; Stayman & Aaker, 1998) solicited measures of ad-evoked feelings and brand evaluations from the same respondents. Such repeated measures raise significant issues of shared method variance, again likely to exaggerate the true effects of ad-evoked feelings on brand evaluations (see Feldman & Lynch, 1988). In summary, given the questionable external validity of most previous studies of the phenomenon, it is difficult to gauge the true magnitude of the effects of ad-evoked feelings on brand evaluations in the real world, with a distinct possibility that these effects may have been overstated.

A second set of issues pertains to the boundary conditions of the phenomenon. In particular, an important question for marketing professionals is whether these effects are generally true across product categories or are instead limited to certain product categories. For example, does an emotional ad have similar effects on attitudes toward a brand of automobiles as opposed to a brand of financial services? Surprisingly enough, this type of question has yet to receive an adequate empirical answer. This is because the relatively small, convenience samples of advertisements—and hence brands and product categories—used in previous studies preclude a rigorous analysis of the generalizability of the phenomenon across product categories. As a result, there is an important gap between what marketing professionals need to know about the scope of the phenomenon and what previous research is able to substantiate.

The purpose of our research is to provide managerially relevant empirical generalizations about the effects of ad-evoked feelings on brand evaluations by (a) addressing key external-validity shortcomings of previous studies, and (b) assessing the generalizability of the phenomenon across product categories. Our investigation departs from previous studies in four major respects. First, instead of soliciting responses from college students, we collected responses from 1576 consumers who were broadly representative of the Belgian population. Second, instead of using a small, convenience sample of advertisements, we used a total of 1070 TV commercials for existing brands, constituting a near census of all commercials shown on Dutch-speaking Belgian television over a three-year period. Third, we used a design that does not require the measurement of ad-evoked feelings and brand evaluations from the same respondents. Finally, we explicitly examined potential product-category-level moderators of the effects of ad-evoked feelings on brand evaluations.

Our results show that even when major issues of external validity are addressed, ad-evoked feelings do have substantial positive effects on brand evaluations, with an effect size that is roughly comparable to that uncovered in previous studies. These effects hold even after controlling for attitude toward the ad (Aad) and cognitive beliefs, suggesting that ad-evoked feelings have direct effects on brand evaluations. The effects appear to be stronger for products typically associated with hedonic motives than for products typically associated with utilitarian motives. However, we found little evidence that these effects depend on whether the product is typically associated with high versus low consumer involvement, whether the product is a durable, a nondurable, or a service, and whether the product is a search or experience good.

2. Conceptual background

2.1. Major theoretical explanations

Four major theoretical explanations can be advanced for the effects of ad-evoked feelings on brand evaluations. The first is that these effects are mediated by changes in consumers’ attitude toward the ad. Ads that evoke more pleasant feelings are typically preferred to ads that evoke less pleasant feelings. Favorable attitudes toward the ad (Aad) translate into favorable attitudes toward the brand (Ab) through a carryover process known as “affect transfer” (MacKenzie, Lutz, & Belch, 1986). Consistent with this explanation, some studies have found that the effects of ad-evoked feelings on brand evaluations are fully mediated by their effects on Aad (e.g., Batra & Ray, 1986; Holbrook & Batra, 1987; Macniss & Park, 1991). Other studies, however, have found only partial mediation (e.g., Burke & Edell, 1989; Stayman & Aaker, 1988), suggesting that additional processes may be at work.

The second explanation posits that these effects are mediated by differences in the beliefs and thoughts that consumers have about the brand. Compared to ads that elicit less pleasant feelings, ads that elicit more pleasant feelings may trigger more positive beliefs and thoughts about the brand, which, when integrated into summary evaluations, would result in more favorable brand attitudes (e.g., Burke & Edell, 1989; Fishbein & Middlestadt, 1995). This second explanation is consistent with research suggesting that immediate feelings toward a target tend to bias subsequent thoughts about the target in the direction of these feelings (e.g., Batra & Stayman, 1990; Pham, Cohen, Pracejus, & Hughes, 2001; Yeung & Wyer, 2004), and with research showing that affective states tend to activate affect-consistent materials in memory (Bower, 1981; Isen, Shalker, Clark, & Karp, 1978). In line with this explanation, some studies have found that the effects of ad-evoked feelings on brand evaluations are largely mediated by changes in brand beliefs and thoughts (e.g., Burke & Edell, 1989; Cho & Stout, 1993; Edell & Burke, 1987), although other studies indicate that the effects of ad-evoked feelings remain substantial after controlling for brand beliefs (Morris et al., 2002).

The third explanation is that the effects are due to a more automatic process of evaluative conditioning (De Houwer, Thomas, & Baeyens, 2001). Specifically, the mere pairing of a brand with the feelings evoked by an ad may result in the valence of these feelings being associatively incorporated into the brand evaluations (Gorn, 1982; Jones, Olson, & Fazio, 2010). This explanation would be consistent with studies indicating direct effects of ad-evoked feelings on brand evaluations after controlling for Aad and brand beliefs (e.g., Burke & Edell, 1989; Cho & Stout, 1993; Homer & Yoon, 1992).

A final explanation is a more inferential process of “affect-as-information” (Schwarz & Clore, 1983, 2007). Given that people often make judgments by inspecting their momentary feelings and asking themselves, “How do I feel about it?” (see Pham, 2004, and Schwarz & Clore, 2007, for reviews), it is possible that consumers interpret their momentary feelings experienced during ad exposure as being indicative of how much they like or dislike the advertised brand. This process is distinct from affect transfer and evaluative conditioning in that the how-do-I-feel-about-it (HDIF) process is inferential, whereas affect transfer and evaluative conditioning are purely associative. The HDIF process is also different from a belief- and thought-priming process in that feelings enter judgments directly, rather than indirectly through a change in beliefs and thoughts.

In summary, ad-evoked feelings can influence brand evaluations through at least four different processes. Given that these processes are probably not mutually exclusive and are instead likely to operate jointly, one would predict that ad-evoked feelings are likely to have substantial positive effects on brand evaluations in the real world, even when major threats to external validity are addressed. Moreover, these effects are likely to be both direct and indirect.
2.2. Potential product-category-level moderators

Given that the purpose of our research is to identify empirical generalizations that are actionable from a managerial standpoint, we focus our analysis of boundary conditions on potential moderators of the phenomenon that are at the product-category level rather than at the individual-consumer level. This is because marketing professionals typically do not know with precision the psychological states of individual consumers, but do know the general characteristics of the product category being advertised. In our study we investigate four potential category-level moderators of the effects of ad-evoked feelings on brand evaluations: (a) whether the product category is typically associated with low or high levels of consumer involvement; (b) whether the product category is typically associated with hedonic versus utilitarian consumption motives; (c) whether the advertised product is a durable, a nondurable, or a service; and (d) whether the product is a “search good,” whose quality can be determined by examination prior to purchase (e.g., curtains, credit cards), or an “experience good,” whose quality can only be determined through actual consumption (e.g., diet programs, pre-prepared meals) (Nelson, 1970).

With respect to a potential moderating role of involvement typically associated with the product category, some studies suggest that the effects of ad-evoked feelings on brand evaluations may be stronger under conditions of lower consumer involvement than under conditions of higher consumer involvement (Batra & Stephens, 1994; Brown & Stayman, 1992; MacInnis & Park, 1991; Madden et al., 1988). If ad-evoked feelings operate as “peripheral cues,” this finding would be consistent with a popular prediction in the persuasion literature that low involvement tends to increase the influence of peripheral heuristic cues on attitudes (Eagly & Chaiken, 1993; Petty & Cacioppo, 1986). However, other considerations would suggest that when assessed at the product-category level (as done in popular planning frameworks) and examined under conditions of greater external validity, involvement may not moderate the effects of ad-evoked feelings on brand evaluations. First, the studies that found increased effects of ad-evoked feelings under low involvement have typically used strong instruction-based manipulations of consumer involvement (e.g., explicit instructions to evaluate the advertised brand vs. instructions to simply watch the ad). It is not clear that “natural” variations in involvement across product categories would be strong enough to produce the type of moderation effects observed in these studies. Second, while certain product categories (e.g., cars) are on average typically more involving than others (e.g., soaps), individual consumers may still vary considerably in their level of involvement with a given product category (e.g., cooking oil for a casual cook versus a professional chef) (Bloch & Richins, 1983).

The considerable heterogeneity across consumers in their level of involvement with a given product category would tend to attenuate differences in typical involvement across product categories. Finally, some research suggests that feelings may influence judgments under conditions of both low involvement and high involvement, albeit through different processes. Whereas under low involvement, ad-evoked feelings may influence brand evaluations through heuristic mechanisms such as affect transfer, evaluative conditioning, or the HDIF heuristic, under high involvement, ad-evoked feelings may influence brand evaluations by shaping the beliefs and thoughts that consumers have about the brand (Albarracin & Wyer, 2001; Batra & Stayman, 1990; Forgas, 1995; Petty, Schumann, Richman, & Strathman, 1993). In other words, involvement may affect the process by which ad-evoked feelings influence brand evaluations rather than the extent to which they influence brand evaluations. In summary, it is not clear whether in real-world settings the effects of ad-evoked feelings on brand evaluations would depend on the level of involvement typically associated with the product category.

With respect to the type of consumption motive, hedonic versus utilitarian, typically associated with the product category, several considerations lead us to predict that it will moderate the effects of ad-evoked feelings on brand evaluations. Specifically, the effects of ad-evoked feelings on brand evaluations are likely to be more pronounced for products that are typically associated with hedonic motives than for products that are typically associated with utilitarian motives. Although this prediction makes intuitive sense and was the rationale for the distinction made in the FCB grid between “think” and “feel” products (Vaughn, 1980, 1986), it has received scant empirical testing in the academic literature. Nevertheless, indirect evidence consistent with this proposition can be found in the literature on incidental mood effects on consumer judgments, which consistently shows that consumers are more influenced by their mood states when they have hedonic motives than when they have utilitarian motives (e.g., Chang & Pham, 2013; Pham, 1998; Pham, Meyvis, & Zhou, 2001). This finding is also observed when comparing the evaluation of hedonic products to the evaluation of utilitarian products (Adaval, 2001; White & McFarland, 2009; Yeung & Wyer, 2004). Therefore, unlike with involvement, we expect to observe this greater influence of feelings among consumers with hedonic as opposed to utilitarian motives extended to settings where these motives are defined at the category level rather than at the individual consumer level. This is because we expect greater homogeneity across consumers in type of motives that they associate with different product categories than in level of involvement with these same product categories.

Two additional product-category characteristics are examined here as potential moderators of the effects of ad-evoked feelings on brand evaluations: whether the product is a durable, a nondurable (e.g., FMCG), or a service; and whether the product is a search good or an experience good. While the literature does not suggest strong a priori predictions about these two product characteristics as moderators of the effects of ad-evoked feelings, they are worth studying on an exploratory basis because they have proven to be important moderators of the impact of advertising in general (Vakratsas & Ambler, 1999; see also Hassens, 2009). Indeed, it has been found that the impact of advertising is greater for experience goods than for search goods, and that this impact may be 50% higher for durable goods than for nondurable goods (see Vakratsas in Hassens, 2009). It has also been found that marketers tend to use different types of advertising for search goods than for experience goods (Nelson, 1970, 1974). It is therefore conceivable that the effects of ad-evoked feelings on brand evaluations might be different depending on these two product-category characteristics.

3. Empirical study

Our study examines (a) how emotional feelings evoked by a large pool of TV commercials influence the brand evaluations of a representative set of adult consumers, and (b) how this influence is moderated by four different product-category-level characteristics. Three important aspects of our investigation should be noted. First, our investigation seeks to better approximate marketplace settings than previous studies generally did. Second, our investigation focuses on the ads themselves, rather than the consumers, as the main units of analysis (as in Holbrook & Batra, 1987). This is because advertisers have greater control over the contents of their ads than over consumers’ responses to these ads. Third, our investigation focuses on product-category-level moderators of the phenomenon, rather than on consumer-level moderators within a category. This is because marketers typically know the general characteristics of the product category but not the specific states of individual consumers. Thus, our research questions are considered at the following level: “Does the emotional pleasantness of ad X influence attitudes toward brand Y, given that Y belongs to product category Z?” which is the way brand managers and advertising executives would contemplate such questions.
3.1. Method

As in Holbrook and Batra (1987), our study relied on an aggregate-covariation design in which the units of analysis were different TV commercials to which the consumers were exposed. A large sample of consumers watched a large number of TV commercials, then reported their attitude toward each advertised brand (Ab) and toward each ad (Aad). The emotions evoked by each commercial were coded by an independent set of judges, and the major characteristics of each product category (involvement, motive, durability, search/experience) were coded by another set of judges. The various responses were averaged across respondents and judges, and the covariation among these responses was modeled across commercials. In addition to being more relevant from a managerial standpoint, this design significantly reduces problems of shared method variance that arise when all the measures are collected from the same respondent (Holbrook & Batra, 1987; see also Pham, Cohen, et al., 2001).

3.1.1. Advertising stimuli

The data were collected in two waves, with the help of a market research firm. For the first wave, we secured a census of all the different brand commercials that aired on a major Dutch-speaking Belgian TV channel over a two-year period. For the second wave, conducted three years later, we secured another census of all brand commercials that aired on a different Dutch-speaking Belgian TV channel over a one-year period. From each census, we excluded commercials directed at children (because respondents were adult consumers), as well as short promotional messages of less than 30 seconds. Ads for “umbrella” brands (e.g., P&G) that were not linked to a specific product category were also excluded. The final stimulus set consisted of a total pool of 1070 commercials (407 for wave 1 and 663 for wave 2), featuring 318 different brands (national and international) across 153 different product categories (e.g., beer, credit cards, diapers, coffee, laundry detergents, cars, computers, etc.). This pool of ads is broadly representative of the full spectrum of brand commercials directed at adult Belgian consumers. The dataset’s summary characteristics are presented in Table 1.

3.1.2. Respondents

Respondents were 1576 Dutch-speaking Belgian consumers who were recruited via TV ads on the same channels and received about €25 for their participation. The recruiting ads generated more than 3000 initial responses for each wave. Of the initial respondents, 1000 were selected in each wave and invited to participate in the study to form a broadly representative sample of adult Belgian consumers. Of those invited, 854 consumers eventually participated in wave 1 and evaluated the first pool of 407 commercials, and 839 consumers participated in wave 2 and evaluated the second pool of 663 commercials. However, due to a variety of factors explained below, only 722 of the 839 wave-2 respondents were retained for the analyses. The demographics of the 1576 eventual consumer respondents are summarized in Table 1.

3.1.3. Procedure and consumer-response measures

The procedure and measures were essentially parallel across the two waves, except for slight differences. In both waves groups of 20 to 30 respondents were invited at regular intervals to a research facility. Each group was shown a subset of the stimulus commercials, one commercial at a time, and asked to rate their responses after viewing each commercial. Two sequences of each subset of ads were used across sessions. Wave-1 respondents were shown about 20 commercials on average \((M = 20.14; SD = 5.64)\), whereas wave-2 respondents were shown about 50 commercials on average, with the number of commercials varying considerably across sessions \((M = 48.73; SD = 23.33)\). (The substantially greater number of commercials shown to wave-2 respondents, which was beyond our control, resulted in lower data quality for the second wave, which necessitated some data purification, as explained below.)

As detailed in Table 2, the consumer respondents completed three main measures for each ad: (a) their attitude toward the ad (Aad), (b) their cognitive assessment of the ad (CogAss) to control for cognitive-belief effects of the ad, and (c) their attitude toward the brand (Ab) as the main dependent measure in this study. Each ad was rated by an average of 43 consumers. These Aad, cognitive assessments (CogAss), and Ab ratings were averaged across respondents to form 1070 aggregate ad-level observations. To verify that the responses were sufficiently homogeneous across respondents, we computed \(\alpha\)-coefficients of inter-respondent agreement for each individual item (Holbrook & Batra, 1987; Pham, Cohen, et al., 2001). As shown in Table 2, the inter-respondent agreement coefficients were high for all items in each of the two waves, thereby justifying an aggregation of the Ab, Aad, and CogAss responses across respondents.

3.1.4. Ad emotional content and creativity

Two independent groups of judges (12 judges for wave 1 and 24 judges for wave 2)—graduate students in marketing who were blind to the study’s hypotheses—rated the emotional content and creativity of each ad (see Table 3 for details). There were six judges per ad. Each wave-1 judge coded half of the wave-1 ads, whereas each wave-2 judge coded one quarter of the wave-2 ads. The ad sequence was rotated across judges. After viewing each ad, the judges first rated the extent to which the ad made them feel various emotions (e.g., excited, sentimental) on a series of 1 (not at all) to 7 (very much) scales, with the order of the items rotated across judges. Given that the vast majority of TV commercials in Belgium are positively valenced, we measured only positive emotions (i.e., warmth-type, excitement-type, and happiness/cheerfulness-type feelings, cf. Burke & Edell, 1989; Edell & Burke, 1987; Holbrook & Batra, 1987). To control for the possibility that ratings of emotional responses to the ads may reflect some other aspects of the ads, such as their originality or creativity, the judges were also asked to rate the ads in terms of creativity.

Table 1

<table>
<thead>
<tr>
<th>Dataset characteristics.</th>
<th>Wave 1</th>
<th>Wave 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># ads</td>
<td>407</td>
<td>663</td>
<td>1070</td>
</tr>
<tr>
<td># products</td>
<td>93</td>
<td>122</td>
<td>215</td>
</tr>
<tr>
<td># brands</td>
<td>197</td>
<td>312</td>
<td>318</td>
</tr>
<tr>
<td># respondents</td>
<td>854</td>
<td>722</td>
<td>1576</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% men</td>
<td>46.7</td>
<td>44.7</td>
<td>45.8</td>
</tr>
<tr>
<td>% women</td>
<td>53.3</td>
<td>55.3</td>
<td>54.2</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% 15–24</td>
<td>46.9</td>
<td>23.5</td>
<td>36.3</td>
</tr>
<tr>
<td>% 25–34</td>
<td>20.0</td>
<td>22.2</td>
<td>21.0</td>
</tr>
<tr>
<td>% 35–44</td>
<td>21.6</td>
<td>34.7</td>
<td>27.5</td>
</tr>
<tr>
<td>% 45–55</td>
<td>11.5</td>
<td>19.6</td>
<td>15.2</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% vocational school degree</td>
<td>40.3</td>
<td>23.7</td>
<td>31.2</td>
</tr>
<tr>
<td>% high school degree</td>
<td>37.8</td>
<td>55.2</td>
<td>47.3</td>
</tr>
<tr>
<td>% technical or community college degree</td>
<td>17.5</td>
<td>12.9</td>
<td>15.4</td>
</tr>
<tr>
<td>% university degree</td>
<td>4.4</td>
<td>8.1</td>
<td>6.2</td>
</tr>
</tbody>
</table>
Hedonic vs. utilitarian product only one set of judges was used.

The notion that attitudes also have a behavioral (conative) component.

Ad-level and product category-level codings.

Table 2
Consumer respondent measures.

<table>
<thead>
<tr>
<th>Wave 1</th>
<th>Wave 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-judge consistency</td>
<td>Inter-item consistency</td>
</tr>
<tr>
<td>Add</td>
<td></td>
</tr>
<tr>
<td>“I like this ad”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = not at all; 10 = very much; Wave 2: −3 = totally disagree, +3 = totally agree)</td>
<td></td>
</tr>
<tr>
<td>“The ad is well made”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = totally disagree; 7 = totally agree; Wave 2: −3 = totally disagree, +3 = totally agree)</td>
<td></td>
</tr>
<tr>
<td>“My general evaluation of the ad is...”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = very negative; 7 = very positive; Wave 2: −3 = very negative, +3 = very positive)</td>
<td></td>
</tr>
<tr>
<td>Cognitive assessment</td>
<td></td>
</tr>
<tr>
<td>“The ad gives useful information”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = totally disagree; 7 = totally agree; Wave 2: −3 = totally disagree, +3 = totally agree)</td>
<td></td>
</tr>
<tr>
<td>“The ad is believable”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = totally disagree; 7 = totally agree; Wave 2: item not measured)</td>
<td></td>
</tr>
<tr>
<td>Ab</td>
<td></td>
</tr>
<tr>
<td>“My evaluation of the brand is...”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = very negative; 7 = very positive; Wave 2: −3 = very negative, +3 = very positive)</td>
<td></td>
</tr>
<tr>
<td>“If I need the product, I would buy this brand”</td>
<td></td>
</tr>
<tr>
<td>(Wave 1: 1 = certainly not; 4 = certainly; Wave 2: −3 = totally disagree, +3 = totally agree)</td>
<td></td>
</tr>
</tbody>
</table>

Inter-judge consistencies

- Inter-respondent agreement (range) = .72–.84
- Inter-item consistency = .74–.84

Inter-item consistencies

- Inter-item reliability (range) = .76–.85
- Mean (SD) = .71–.81

Ad emotional content

Warmth

- “Sentimental” .76–.78 .69–.77
- “Emotional” .75–.78 .67–.75
- “Moves me” .78–.81 .69–.75
- “Warm-hearted” .76–.76
- “Touched me” .73–.75

Excitement

- “Energetic” .61–.79 .69–.79
- “Excited” .72–.77 .68–.78
- “Enthusiastic” .65–.74 .68–.79
- “Upbeat” .69–.75
- “Stimulated” .66–.77

Happiness/cheerfulness

- “Cheerful” .74–.83 .70–.79
- “Joyful” .72–.82 .71–.81
- “In a good mood” .72–.84 .71–.81
- “Happy” .75–.82
- “Delighted” .74–.84

Ad creativity

- “Unique” .80–.83 .75–.87
- “Original” .77–.86 .82–.85
- “Creative” .81–.82 .78–.86
- “Novel” .79–.84
- “Unlike other ads” .83–.84

Product involvement

- “[x] means a lot/nothing to consumers” .72 .82
- “[x] is of great/little concern to consumers” .68 .80
- “Choosing [x] is/is not an important decision” .80 .90
- “There is substantial/no risk involved with [x]” .74 .83
- “It is/is not a big deal if consumers make a mistake when buying [x]” .78 .89

Hedonic vs. utilitarian product

- “[x] is more a luxury than a necessity/more a necessity than a luxury” .81 .93
- “The benefits are primarily hedonic/functional” .87 .92
- “With [x] sensations and sensory stimulations play an important/Minor role” .81 .86
- “The motivation for using [x] is mainly emotional/rational” .87 .93
- “This is a product/service consumers use for pleasure or to impress others/to address or avoid problems” .86 .92

Although this item can be seen as a measure of behavioral intention, we treat it as a measure of brand attitude because of its high correlation with the other Ab item and the notion that attitudes also have a behavioral (conative) component.}

Table 3
Ad-level and product category-level codings.

<table>
<thead>
<tr>
<th>Wave 1</th>
<th>Wave 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-judge consistencies</td>
<td>Inter-item consistencies</td>
</tr>
<tr>
<td>Ad emotional content</td>
<td></td>
</tr>
<tr>
<td>Warmth</td>
<td></td>
</tr>
<tr>
<td>“Sentimental”</td>
<td>.97</td>
</tr>
<tr>
<td>“Emotional”</td>
<td></td>
</tr>
<tr>
<td>“Moves me”</td>
<td>.76–.78</td>
</tr>
<tr>
<td>“Warm-hearted”</td>
<td>.75–.78</td>
</tr>
<tr>
<td>“Touched me”</td>
<td>.78–.81</td>
</tr>
<tr>
<td>Happiness/cheerfulness</td>
<td></td>
</tr>
<tr>
<td>“Cheerful”</td>
<td>.74–.83</td>
</tr>
<tr>
<td>“Joyful”</td>
<td>.72–.82</td>
</tr>
<tr>
<td>“In a good mood”</td>
<td>.72–.84</td>
</tr>
<tr>
<td>“Happy”</td>
<td></td>
</tr>
<tr>
<td>“Delighted”</td>
<td></td>
</tr>
<tr>
<td>Ad creativity</td>
<td></td>
</tr>
<tr>
<td>“Unique”</td>
<td>.80–.83</td>
</tr>
<tr>
<td>“Original”</td>
<td>.77–.86</td>
</tr>
<tr>
<td>“Creative”</td>
<td></td>
</tr>
<tr>
<td>“Novel”</td>
<td>.79–.84</td>
</tr>
<tr>
<td>“Unlike other ads”</td>
<td>.83–.84</td>
</tr>
<tr>
<td>Product involvement</td>
<td></td>
</tr>
<tr>
<td>“[x] means a lot/nothing to consumers”</td>
<td>.72</td>
</tr>
<tr>
<td>“[x] is of great/little concern to consumers”</td>
<td>.68</td>
</tr>
<tr>
<td>“Choosing [x] is/is not an important decision”</td>
<td>.80</td>
</tr>
<tr>
<td>“There is substantial/no risk involved with [x]”</td>
<td>.74</td>
</tr>
<tr>
<td>“It is/is not a big deal if consumers make a mistake when buying [x]”</td>
<td>.78</td>
</tr>
<tr>
<td>Hedonic vs. utilitarian product</td>
<td></td>
</tr>
<tr>
<td>“[x] is more a luxury than a necessity/more a necessity than a luxury”</td>
<td>.81</td>
</tr>
<tr>
<td>“The benefits are primarily hedonic/functional”</td>
<td>.87</td>
</tr>
<tr>
<td>“With [x] sensations and sensory stimulations play an important/Minor role”</td>
<td>.81</td>
</tr>
<tr>
<td>“The motivation for using [x] is mainly emotional/rational”</td>
<td>.87</td>
</tr>
<tr>
<td>“This is a product/service consumers use for pleasure or to impress others/to address or avoid problems”</td>
<td>.86</td>
</tr>
</tbody>
</table>

Entries represent the range of inter-judge consistencies across the different sets of judges for the measures of Ad emotional content and Ad creativity. For Product involvement and Hedonic vs. utilitarian product only one set of judges was used.

In wave 2, only three of the five items used in wave 1 were used to assess Ad emotional content and Ad creativity.
across judges were also internally consistent across items and were therefore averaged to form a single score of creativity for each ad.

3.1.5. Product-category-level involvement and motivation type

To test potential product-category-level moderators of the effects of ad-evoked emotional feelings, two other sets of six judges (who were also blind to the hypotheses) coded the 153 product categories represented in the ads. The first set of judges coded the 93 product categories featured in the wave-1 ads, whereas the second set of judges coded an additional 60 categories featured in wave-2 ads that were not included in the wave-1 ads. Unlike the other judges, these judges did not watch the ads. Instead, they were simply given the names of the product/service categories (e.g., “cell phones,” “employment agency,” “pizza,” “paper towels”) and asked to rate each category on five semantic differential items assessing product involvement (based on Laurent & Kapferer, 1985; Zaichkowsky, 1985) and five semantic differential items assessing hedonic versus utilitarian motives (adapted from Dhar & Wertenbroch, 2000; Voss, Spangenberg, & Grohmann, 2003) (see Table 3). The order of both the product category names and the coding items was rotated across judges. Inter-judge agreement was \( \alpha = .78 \) for both items, again justifying an aggregation.

3.1.6. Durable–nondurable–service and search–experience product categorization

Finally, another set of four judges was asked to categorize each of the 153 product categories into (1) “a durable product, that is, a tangible product that lasts some time, such as refrigerators,” (2) “a nondurable product, that is, a tangible product that usually is fully consumed in one or a few uses, such as beer or soap,” or (3) “a service, that is, an intangible, inseparable, variable, and short-term provided product such as a haircut or a phone subscription” (inter-judge agreement = 97.4%; disagreements resolved by majority rule). The same four judges additionally rated the extent to which each of the 153 product categories was a search good or an experience good using two 5-point semantic differential scales inspired by Nelson (1970, 1974): (1) “This is a product for which it is easy for the consumer to evaluate the quality before the purchase by inspection (e.g., clothing, cookware, luggage, furniture, etc.)” and (2) “The quality of this product or service is pretty obvious even without trying it” versus “The quality of this product or service can only be determined by trying it and experiencing it.” The order of both the product category names and the coding items was rotated across judges. Inter-judge agreement was \( \alpha = .78 \) for both items, again justifying an aggregation.

3.1.7. Data purification of wave 2

Although the overall pattern of data was generally consistent across data-collection waves, preliminary analyses indicated a substantially greater level of noise in the second wave. We attribute this greater noise to the fact that respondents in the second wave had to watch and evaluate on average more than twice as many commercials as respondents in the first wave (for example, many wave-2 respondents evaluated between 75 and 101 commercials). Because different respondents saw a different number of ads and because the order in which the ads were seen by individual respondents was not available to us, it was not possible to restrict our analyses to the first 20 or so advertisements that each wave-2 respondent evaluated. Therefore, in order to make the wave-2 data more comparable in terms of quality with the wave-1 data, the former were purified as follows. First, we eliminated any responses to a given ad by a given respondent that indicated a clear lack of care through either two or more missing values across items or identical responses across all items (e.g., 5, 5, 5, 5…). This resulted in the elimination of 8325 out of 41,320 total observations (20.2%). Next, we estimated whether respondents were diligent in their responses or responded quasi-randomly by fitting, for each respondent, a regression model in which the respondent’s attitude toward a given brand was to be explained by the respondent’s attitude toward the ad (Aad) and the respondent’s brand familiarity. Any respondent whose regression R² was less than .10 was considered to be a quasi-random responder and thus dropped from the data. This resulted in the further elimination of 2794 observations (6.8%) from 77 respondents. Finally, we restricted the data to those respondents who showed a strong degree of internal consistency of .80 in Aad responses across advertisements. This resulted in the further elimination of 1522 observations (3.7%). The purified wave-2 data set thus consists of 28,679 observations (69.4%) from 722 respondents evaluating a total of 663 commercials.  

Table 4

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Correlations (wave 1, n = 407)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ab</td>
</tr>
<tr>
<td>Ab</td>
<td>1</td>
</tr>
<tr>
<td>Aad</td>
<td>1</td>
</tr>
<tr>
<td>Ad emotional content</td>
<td>1</td>
</tr>
<tr>
<td>Cognitive assessment</td>
<td>1</td>
</tr>
<tr>
<td>Ad creativity</td>
<td>1</td>
</tr>
<tr>
<td>Product involvement</td>
<td>1</td>
</tr>
<tr>
<td>Hedonic vs. utilitarian</td>
<td>1</td>
</tr>
</tbody>
</table>

***p < .001; **p < .01; *p < .05 (two-tailed significance tests).

1. A higher score indicates more positive emotions.
2. A higher score indicates more hedonic motives.
3. A higher score indicates relatively more experience goods.

1 If all wave-2 respondents are retained in the analyses, the results are directionally the same as reported in Tables 6 and 7, but are statistically weaker. The main difference in the results is that the emotion × hedonic/utilitarian motive interaction does not reach significance in Models 7 and 8 using hierarchical linear model regression, although it does reach significance in OLS regression.
3.2. Results

Tables 4 and 5 report descriptive statistics for all the variables, along with the simple correlations among them. Because a significant proportion of the brands had more than one commercial and because different brands would often share the same product category (e.g., Nissan and Mercedes), the data were analyzed in a series of multilevel regression models in which emotional responses and their potential mediators and moderators were treated as fixed effects, and brand and product-category effects were modeled as random intercepts with the brand effects nested in product category. These three-level models, with ads nested in brands and brands nested in product categories, adjust for any data dependencies that may exist across ads for the same brands and across brands within the same product category (simpler OLS regression models produced largely similar results). All predictors in the models were mean-centered, with the effects of ad-evoked feelings, involvement, and hedonic versus utilitarian motives being additionally standardized as factor loadings.

3.2.1. Effects of ad-evoked feelings

Table 6 summarizes the results of four models testing the basic effects of ad-evoked feelings on brand evaluations and the potential mediators of these effects. The results of Model 1, the “basic-feeling-effect model,” show that even under conditions of greater external validity, ad-evoked feelings indeed have a substantial influence on brand evaluations ($r = .486, t = 13.69, p < .001$). Interestingly, the simple correlation between ad-evoked feelings and brand evaluations was $r = .331$, which is roughly of the same magnitude as what had been

| Table 5 |
| Correlations (wave 2, n = 663). |

<table>
<thead>
<tr>
<th></th>
<th>Ab</th>
<th>Aad</th>
<th>Ad emotional content</th>
<th>Cognitive assessment</th>
<th>Ad creativity</th>
<th>Product involvement</th>
<th>Hedonic vs utilitarian</th>
<th>Durability or service (0) vs. (Non-)Durable (0) vs Service (1)</th>
<th>Search vs. experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab</td>
<td>1</td>
<td>.501***</td>
<td>.256***</td>
<td>.316***</td>
<td>.116***</td>
<td>-.351***</td>
<td>-.005</td>
<td>.262***</td>
<td>-.227***</td>
</tr>
<tr>
<td>Aad</td>
<td>1</td>
<td>.496***</td>
<td>.130***</td>
<td>.441***</td>
<td>-.015</td>
<td>.090***</td>
<td>-.032</td>
<td>.090***</td>
<td>-.091***</td>
</tr>
<tr>
<td>Ad emotional content</td>
<td>1</td>
<td>-.162***</td>
<td>.584***</td>
<td>-.056</td>
<td>.120***</td>
<td>-.091***</td>
<td>.058</td>
<td>-.111***</td>
<td>-.129***</td>
</tr>
<tr>
<td>Cognitive assessment</td>
<td>1</td>
<td>-.324***</td>
<td>-.052</td>
<td>-.344***</td>
<td>-.155***</td>
<td>-.333***</td>
<td>-.245***</td>
<td>-.312***</td>
<td>-.312***</td>
</tr>
<tr>
<td>Ad creativity</td>
<td>1</td>
<td>-.086</td>
<td>.011</td>
<td>-.333***</td>
<td>-.245***</td>
<td>-.312***</td>
<td>-.387***</td>
<td>.011</td>
<td>-.333***</td>
</tr>
<tr>
<td>Product involvement</td>
<td>1</td>
<td>.000</td>
<td>-.574***</td>
<td>.111</td>
<td>-.220***</td>
<td>-.261***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
</tr>
<tr>
<td>Hedonic vs. utilitarian</td>
<td>1</td>
<td>-.220***</td>
<td>-.261***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
</tr>
<tr>
<td>Durability/service (0) vs. (Non-)Durable (0) vs Service (1)</td>
<td>1</td>
<td>.000</td>
<td>-.574***</td>
<td>.111</td>
<td>-.220***</td>
<td>-.261***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
</tr>
<tr>
<td>Search vs. experience</td>
<td>1</td>
<td>.000</td>
<td>-.574***</td>
<td>.111</td>
<td>-.220***</td>
<td>-.261***</td>
<td>-.241***</td>
<td>-.241***</td>
<td>-.241***</td>
</tr>
</tbody>
</table>

**p < .001; *p < .01; *p < .05 (two-tailed significance tests).

Table 6

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic-feeling-effect model</td>
<td>Aad-mediation model</td>
<td>Cognitive-assessment-and-Aad mediation model</td>
<td>Ad-creativity-confound model</td>
</tr>
<tr>
<td>Basic fit</td>
<td>AIC 3073.9</td>
<td>2781.8</td>
<td>2712.3</td>
</tr>
<tr>
<td></td>
<td>BIC 3083.0</td>
<td>2790.9</td>
<td>2721.4</td>
</tr>
<tr>
<td></td>
<td>Chi$^2$ 437.19***</td>
<td>493.58***</td>
<td>459.63***</td>
</tr>
<tr>
<td>Random effects</td>
<td>Product .750***</td>
<td>.523***</td>
<td>.424***</td>
</tr>
<tr>
<td></td>
<td>Brand .557***</td>
<td>.486*</td>
<td>.466***</td>
</tr>
<tr>
<td></td>
<td>(8.88)</td>
<td>(10.23)</td>
<td>(10.53)</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>Emotional content .486***</td>
<td>.108*</td>
<td>.162***</td>
</tr>
<tr>
<td></td>
<td>Wave (.72)</td>
<td>(2.81)</td>
<td>(4.31)</td>
</tr>
<tr>
<td></td>
<td>Emotional content $^a$ $^b$ × Wave $^b$ .279***</td>
<td>.080*</td>
<td>.113**</td>
</tr>
<tr>
<td></td>
<td>Aad .692***</td>
<td>(2.11)</td>
<td>(3.04)</td>
</tr>
<tr>
<td></td>
<td>(18.07)</td>
<td>(9.46)</td>
<td>(10.03)</td>
</tr>
<tr>
<td></td>
<td>Aad × Wave $^b$ .216***</td>
<td>.022</td>
<td>.097*</td>
</tr>
<tr>
<td></td>
<td>(5.71)</td>
<td>(5.0)</td>
<td>(2.14)</td>
</tr>
<tr>
<td></td>
<td>Cognitive assessment $^c$ .599***</td>
<td>.445***</td>
<td>.510***</td>
</tr>
<tr>
<td></td>
<td>(8.83)</td>
<td>(9.46)</td>
<td>(10.03)</td>
</tr>
<tr>
<td></td>
<td>Cognitive assessment $^c$ × Wave $^b$ .235***</td>
<td>(3.71)</td>
<td>(7.90)</td>
</tr>
<tr>
<td></td>
<td>Ad creativity $^c$ $^b$</td>
<td>.081*</td>
<td>.150</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(2.14)</td>
<td>(2.14)</td>
</tr>
<tr>
<td></td>
<td>Ad creativity $^c$ $^b$ × Wave $^b$</td>
<td>.081*</td>
<td>.150</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(2.14)</td>
<td>(2.14)</td>
</tr>
</tbody>
</table>

Note. **p < .001; *p < .01; *p < .05 (two-tailed significance tests).

$^a$ A higher score indicates more positive emotional feelings.

$^b$ Wave $^b$ = 1; Wave $^b$ = −1.
Table 7
Standardized regression coefficients and z-values (for random effects) and t-values (for fixed effects) \((n = 1070\) ads).

<table>
<thead>
<tr>
<th></th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Involvement-moderation model</td>
<td>Hedonic/utilitarian-moderation model</td>
<td>Involvement-and-hedonic/utilitarian-moderation model</td>
<td>Product-durability-moderation model</td>
<td>Search/experience-moderation model</td>
<td>Full model</td>
</tr>
<tr>
<td><strong>AIC</strong></td>
<td>2630.6</td>
<td>2682.1</td>
<td>2637.2</td>
<td>2646.7</td>
<td>2665.7</td>
<td>2634.1</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>326.3***</td>
<td>432.70***</td>
<td>310.09***</td>
<td>344.37***</td>
<td>290.87***</td>
<td></td>
</tr>
<tr>
<td><strong>Chi²</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hedonic/utilitarian-moderation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cognitive assessment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ad creativity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Product involvement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Emotional content</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aad</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wave</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. ***p < .001; **p < .01; *p < .05; +p < .10 (two-tailed significance tests).

A higher score indicates more positive feelings.

A higher score indicates more hedonic motives.

A higher score indicates relatively more experience goods.

Wave \(1 = 1; \) wave \(2 = -1\). Although not included in the table, all interactions with wave were also modeled.
observed in a meta-analysis of previous studies (Brown et al., 1998). Therefore, it appears that the basic effects are of genuinely substantial size and are not driven by methodological artifacts such as the use of student respondents, a selection bias in the ads used, and shared method variance due to the repeated measurement of respondents. However, a significant interaction with wave ($\beta = -0.379, t = 8.16, p < .001$) indicated that the effects of ad-evoked feelings were substantially stronger in wave 1 than they were in wave 2 (see also Tables 4 and 5). The difference between the two waves could be due to a genuine difference in the strength of the effects depending on the pool of advertisements studied or to methodological differences between the two waves (e.g., in wave 2 respondents evaluated a much larger number of ads and ad-evoked feelings were assessed with fewer items). (As reported in Table 6, several other interactions with wave were uncovered in models 2–4. These interactions are not discussed here because of their lower theoretical and substantive importance.)

The results of Model 2, the “Aad-mediation model,” show that respondents’ attitudes toward the ads (Aad) are strong predictors of their attitudes toward the brands (Ab) ($\beta = .692, t = 18.07, p < .001$) (cf. McKenzie et al., 1986; Mitchell & Olson, 1981). The results additionally show that the effect of ad-evoked feelings is substantially reduced when respondents’ attitudes toward the ads are controlled for ($\beta = .486 \rightarrow \beta = .108$), though the effect remains significant ($t = 2.81, p = .005$). Thus, according to Alwin and Hauser’s (1975) simple formula, as much as 78% of the effects of ad-evoked feelings on brand evaluations ($t = 2.49, p < .001$) may be mediated by changes in Aad. This result is consistent with the notion that effects of ad-evoked feelings on brand evaluations are largely mediated by their effects on ad attitudes, as suggested by various authors (e.g., Batra & Ray, 1986; Holbrook & Batra, 1987; Maclnnis & Park, 1991), but the mediation is not complete, as suggested by other authors (e.g., Burke & Edell, 1989; Stayman & Aaker, 1988).

The results of Model 3, the “cognitive-assessment and Aad-mediation model,” show that while respondents’ cognitive assessments of the ads influenced their brand evaluations ($\beta = .599, t = 8.83, p < .001$), these assessments did not attenuate the effects of ad-evoked feelings on brand evaluations ($\beta = .162, t = 4.31, p < .001$). Although these cognitive assessments of the ads are not actual measures of brand beliefs, this finding seems somewhat inconsistent with a pure belief-based explanation of the phenomenon.

Finally, the results of Model 4, the “ad-creativity-confound model,” show that while the direct effects of ad-evoked feelings on brand evaluations remain ($\beta = .199, t = 4.83, p < .001$) even after controlling for differences in ad creativity ($\beta = -0.081, t = -2.32, p = .02$). This finding suggests that the observed effects of ad-evoked feelings are not confounded by executional elements of the ads such as their creativity.

Overall, the results of these initial analyses support two empirical generalizations. First, even under conditions of greater external validity, ad-evoked feelings exert a substantial positive influence on brand evaluations (EG1). Second, ad-evoked feelings have both direct and indirect effects on brand evaluations, with the indirect effects being stronger and largely linked to changes in Aad (EG2).

### 3.2.2. Product-category moderators of the effects of ad-evoked feelings

Table 7 summarizes the results of five models testing the potential product-category-level moderators of the effects of ad-evoked feelings on brand evaluations. The results of Model 5, the “involvement-moderation model,” indicate that while category-level involvement has a “main” effect on brand evaluations—brand evaluations were less favorable for high-involvement products than for low-involvement products ($\beta = -0.408, t = -7.69, p < .001$)—it did not moderate the effects of ad-evoked feelings on brand evaluations ($\beta = -0.021, t = -0.82, p = .414$).

The results of Model 6, the “hedonic/utilitarian-moderation model,” indicate that the effects of ad-evoked feelings on brand evaluations were marginally stronger for product categories typically associated with hedonic motives than for product categories typically associated with utilitarian motives ($\beta = .054, t = 1.75, p = .081$). This finding was not qualified by an interaction with wave ($\beta = .033, t = 1.09, p = .275$).

Because popular planning models such as the FCB grid and the Rossiter–Percy grid (Rossiter, Percy, & Donovan, 1991) generally conceptualize the advertising effects of product involvement and product motive (hedonic vs. utilitarian) in a two-dimensional space, Model 7, the “involvement- and hedonic/utilitarian-moderation model,” tests jointly for the moderating effects of category-level involvement and type of motive. The results again indicate that product-category-level involvement does not moderate the effects of ad-evoked feelings on brand evaluations ($\beta = -0.041, t = -1.49, p = .137$). However, the type of motive does moderate the effects of ad-evoked feelings on brand evaluations, with the effects being stronger for product categories typically associated with hedonic motives than for product categories typically associated with utilitarian motives ($\beta = -0.078, t = 2.49, p = .013$) (see Fig. 1). Spotlight analyses (Aiken & West, 1996) show that the effect (slope) of ad-evoked feelings was significant for the relatively more hedonic product categories (slope = .350, SE = .053, $t = 6.57, p < .001$), but only marginally significant for the relatively more utilitarian product categories (slope = .098, SE = .056, $t = 1.74, p = .082$). Again, the effects were not qualified by an interaction with wave ($\beta = .049, t = 1.59, p = .112$). (There was no three-way interaction among feelings, involvement, and type of motive ($\beta = -0.038, t = -1.10, p = .231$).)

Overall, the results of Models 5 through 7 support two additional empirical generalizations. First, the effects of ad-evoked feelings on

![Fig. 1. Emotional content x Type of motive interaction on brand attitude.](image-url)
brand evaluations do not appear to depend on the level of involvement typically associated with the product category (EG3). Second, the effects of ad-evoked feelings on brand evaluations are more pronounced for products typically associated with hedonic motives than for products typically associated with utilitarian motives (EG4).

The results of Model 8, the “product-durability-moderation model,” indicate that while brand evaluations were more favorable for nondurable products than for durable products ($\beta = .601$, $t = 3.15$, $p = .002$), the effects of ad-evoked feelings on brand evaluations did not depend on whether the product was a durable, a nondurable, or a service ($\beta = .111$, $t = 1.21$, $p = .226$; $\beta = .025$, $t = 0.25$, $p = .805$). Similarly, the results of Model 9, “the search/experience-moderation model,” indicate that while brand evaluations were more favorable for experience products than for search products ($\beta = .250$, $t = 3.40$, $p = .001$), the effects of ad-evoked feelings on brand evaluations did not depend on whether the product was a search good or an experience good ($\beta = .047$, $t = 1.30$, $p = .195$). Thus, even though the two general product characteristics have been found to be important moderators of the overall effectiveness of advertising (Vakratsas in Hansens, 2009), product durability and the search-versus-experience nature of the good do not appear to moderate the effects of ad-evoked feelings on brand evaluations (EGS).2

The results of the final model, Model 10 (the “full model”), support the inferences suggested by the more restricted models. Category-level involvement did not moderate the effects of ad-evoked feelings on brand evaluations ($\beta = -.043$, $t = -1.39$, $p = .165$), but the effects of ad-evoked feelings were marginally stronger for hedonic products than for utilitarian products ($\beta = .080$, $t = 1.72$, $p = .085$). (The three-way interaction among ad-evoked feelings, involvement, and motive was not significant; $p = .135$.) The interactions between ad-evoked feelings and the product-type dummies were not significant ($t's < 1$), suggesting that the effects of ad-evoked feelings did not depend on whether the product was a durable, a nondurable, or a service. Finally, there was no interaction between ad-evoked feelings and whether the product was a search good or an experience good ($t < 1$).

4. General discussion

Considering the practical significance of the effects of ad-evoked feelings on brand evaluations, it is somewhat surprising that the empirical generalizability of this phenomenon—in terms of both external validity and boundary conditions across product categories—had yet to be systematically investigated. Our study addresses this void by analyzing consumer responses to a total of 1070 brand TV commercials from more than 150 different product categories. Unlike previous studies that often involved student respondents, limited samples of often fictitious ads, and repeated measurement of respondents, our study (a) examined the evaluation responses of a large and broadly representative sample of actual consumers, (b) was based on a virtual census of all ads (for real brands) shown by two national TV channels during a three-year period, and (c) used a design that reduces issues of shared method variance.

The results show that even under conditions that are closer to marketplace settings than most previous academic studies, ad-evoked feelings indeed have a substantial impact on brand evaluations (EG1). The effect size was in fact quite comparable to that found in a meta-analysis of these earlier studies (Brown et al., 1998). The results additionally show that the effects of ad-evoked feelings on brand evaluations are both direct and indirect, with the indirect effects being substantially stronger and largely linked to a change in Aad (EG2). The finding that ad-evoked feelings have both direct and indirect effects helps reconcile previously conflicting results that documented either one or the other.

Important additional results pertain to the category-level moderators of the phenomenon. First, we found little evidence of a moderating role of product-category-level involvement (EG3). Given the very large number of observations in our study and the recorded reliability of the measures, we believe that this lack of moderating effect of category-level involvement is more than an artifact of poor measurement or low statistical power. Rather, it is a substantive and generalizable result. We suspect that natural variations in consumer involvement as a function of product category are not very strong in the real world due to substantial inter-consumer heterogeneity in involvement within a product category. In addition, it is possible that feelings can influence judgments under conditions of both low involvement and high involvement, albeit through different mechanisms.

Second, while the level of involvement with the product category does not appear to significantly moderate the effects of ad-induced feelings on brand evaluations, the type of motive typically associated with the category does. The effects of ad-induced feelings on brand evaluations appear to be significantly more pronounced when the product category is more hedonic than when it is more utilitarian (EG4). This is consistent with research in the affect-as-information literature indicating that consumers are more likely to rely on their momentary feelings in judgments and decisions when they have experiential motives than when they have instrumental motives (Pham, 1998; see also Adaval, 2001; Yeung & Wyer, 2004).

Although our study design does not allow for strong process inferences, this finding has potential theoretical implications for research on both the basic phenomenon and the affect-as-information framework. With respect to the former, our findings suggest that the robust effects of ad-evoked feelings on brand evaluations may not be driven solely by affect transfer, evaluative conditioning, and differences in brand beliefs and thoughts, as previously suggested. The phenomenon may also be driven by HDIF inferences during ad exposure, whereby consumers may interpret their feeling responses to the ad as indicative of how much they like or dislike the brand. With respect to the affect-as-information literature, it is important to note that in our studies consumers who were not explicitly asked to assess their feelings toward the various brands nevertheless appeared to incorporate ad-evoked feelings selectively as a function of the type of motive typically associated with the advertised product category. This finding may suggest that the selective reliance on feelings as a function of their relevance for the judgment at hand is a very spontaneous and flexible process that is much more flexible than conceptualized in early affect-as-information research (Schwarz & Clore, 1983).

Finally, while product durability and the search-versus-experience nature of the good have been found to be important moderators of the overall impact of advertising, we found no evidence that these two general product characteristics play any role in moderating the effects of ad-evoked feelings on brand evaluations (EG5).

One obvious limitation of the study is that because the units of analysis were at the aggregate ad level, rather than at the individual respondent level, the results do not allow strong theoretical inferences about the actual psychological processes at work. In addition, it would have been useful to control for prior brand attitude, which...
we were unable to do with these data. Moreover, given evidence of qualitative difference among distinct emotions (e.g., Batra & Ray, 1986; Raghunathan, Pham, & Corfman, 2006), it would also have been interesting to distinguish among subtypes of ad-evoked feelings, which was not possible here because the subtypes of feelings that we assessed were too correlated.

Our findings have obvious managerial implications. Advertisers do benefit substantially from advertisements that elicit pleasant emotional feelings, not only in terms of greater ad liking, but more importantly in terms of more favorable brand attitudes. The total effects of ad-evoked feelings on brand evaluations are substantial. In addition, they appear to be rather general across product categories. They apply equally to (a) low- and high-involvement products, (b) durable products, nondonurable products, and services, and (c) search and experience products. However, the effects are more likely to occur in product categories typically associated with hedonic motives than in product categories typically associated with utilitarian motives. While one could think that the latter proposition should be obvious to advertisers, there is evidence that it is not. For example, it has been found that although food products are typically associated with hedonic and experiential motives in consumers’ minds, food advertisers still tend to rely on informational appeals rather than more emotional ones (Dube et al., 1996). Hence the importance of revisiting what is assumed to be known, even after 25 years.

References


How contracts and enforcement explain transaction outcomes

Erik A. Mooi, David I. Gilliland

A R T I C L E   I N F O
Article history:
First received in 16, April 2012 and was under review for 6 months
Available online 20 June 2013
Guest Editor: Marnik G. Dekimpe
Area Editor: Stefan Wuyts

Keywords:
Enforcement
Contracts
Business marketing
Governance

A B S T R A C T
This study considers the influence of contracts on enforcement and the subsequent performance impact of aligned and misaligned enforcement. We define enforcement as a corrective action aimed at remedying problems occurring in the transaction. First we explain the role of contracts and show that at the component level, contracts can both increase and decrease enforcement. Building on an alignment perspective and accounting for the endogeneity of enforcement, we use these contractual components and variables related to enforcement to predict the occurrence of enforcement. We use such predictions to show that aligned enforcement results in higher performance. We also show that the performance impact of misaligned enforcement is relatively greater for transactions where enforcement is not expected. We conduct the study using a unique dataset reporting on 971 business transactions across a wide range of industries.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Enforcement is a corrective action aimed at remedying problems (Antia et al., 2006). Taking such corrective action requires firms to balance the benefits of enforcement against its costs. The key benefit is that it may curb or reverse violations of contractual agreements (Antia et al., 2006). Enforcement may also reduce or reverse behaviors such as suppliers not remediing product breakdowns or providing limited or inadequate service. Thus, enforcement may help suppliers resolve problems (Wuyts, 2007). On the other hand, exchange partners stung by enforcement may react through further acts, such as protracted conflicts, retaliation, or even relationship termination (Antia & Frazier, 2001). As such, firms need to understand when to enforce and the potential consequences of enforcement.

The role of contracts on enforcement is little understood. One perspective is that having explicit contractual agreements ex ante can facilitate or even trigger enforcement ex post. In fact prior literature, predominantly taking an agency perspective, assumes that enforcement is automatically triggered when contracts are violated (cf. Bergen, Heide, & Dutta, 1998). Another perspective is that such explicit contractual agreements may reduce transaction problems or promote cooperation, thereby reducing the need for enforcement (Mooi & Ghosh, 2010). Recent work acknowledges the role of contracts in enforcement but has conceptualized contracts as monolithic governance devices (cf. Kashyap, Antia, & Frazier, 2012). A more fine-grained analysis of the effects of contracts on enforcement is needed to advance our understanding of whether and how contracts impact enforcement.

We also know little about the performance consequences of enforcement. Recent work correlated enforcement with outcomes but found no effects (Kashyap et al., 2012). Taking a discriminating alignment position may help uncover performance consequences as enforcement is likely best used when matched to circumstances. Based on governance theories, such as transaction cost economics (TCE), the discriminating alignment view argues that governance (enforcement in our case) that is aligned (expected or called for, as based on transactional attributes) may help performance while misaligned enforcement is detrimental to performance. Such an alignment approach to enforcement has, however, not been examined conceptually and empirically. Moreover, the performance implications of misalignment are not well understood. Specifically, comparing the differential performance of aligned enforcement with misaligned enforcement provides insight into the cost of mistakes. Such analyses are rare, yet valuable, as they provide evidence of the importance of carefully choosing governance (Masten, 1993).

The goal of this paper is to study the effects of contracts on enforcement, to understand the performance effects of aligned enforcement, and to understand the performance consequences of misaligned enforcement. In doing so we make three contributions.

Our first contribution is to describe the role of contracts in enforcement. In this light, Bergen et al. (1998) suggest that the assumption is often made that once contracts are in place, the ex post management task is trivial. We demonstrate that enforcement is not automatic and different contractual components can both increase and decrease the use of enforcement. As such, we also show that contracts are not...
monolithic governance structures. To support this contribution, we argue that terms in the contract that support the parties’ relationship (e.g., joint management, nondisclosure) reduce enforcement, while terms designed to protect the transaction increase enforcement.

Our second contribution is to test the importance of alignment between these contractual components, transactional attributes, and enforcement. By comparing the outcomes of aligned (predicted) versus nonaligned (not predicted) enforcement, we account for the little-researched issue of the benefits of aligned governance in an enforcement context (Geyskens, Steenkamp, & Kumar, 2006). We consider performance consequences in terms of satisfaction with problem resolution, which is the satisfaction of the buyer with how problems regarding the product have been resolved. Satisfaction is fundamental to understanding interfirm relationships (Geyskens & Steenkamp, 1999).

A third related contribution is to provide understanding of the performance consequences of misalignment. Prior work has found interesting asymmetries regarding the circumstances under which misalignment has the most severe (negative) performance implications (Ghosh & John, 2009). Specifically, such work suggests that under greater hazards, misalignment has the most severe consequences. Addressing this issue in an enforcement context helps us understand where the risks are in making enforcement choices and helps managers make informed decisions.

We conduct our investigations by using the External Management of Automation dataset, access to which is provided by the Steinmetz Archive. This unique dataset reports in detail on 971 randomly selected transactions executed between information technology (IT) buyers and suppliers. It includes a broad spectrum of firms from industries such as logistics, parts production, and wholesaling.

This paper proceeds by discussing theory on enforcement in Section 2. We develop arguments on the structure of contracts and the expected effects of different contractual components on enforcement in Section 2.1. We continue by building hypotheses on why aligned enforcement results in better performance in Section 2.2. In Section 2.3 we argue that relative performance loss is higher when buying firms mistakenly enforce.

2. Theory and hypotheses

Enforcement is an important governance mechanism in economics, contract law, and marketing (Crocker & Masten, 1991; Williamson, 1996). Despite the importance of enforcement, little work in marketing considers enforcement. Exceptions include Dutta, Bergen, and John (1994), Antia and Frazier (2001), Gilliland and Bello (2002), and Kashyap et al. (2012). Despite these efforts, the role of contracts in enforcement is little understood, as are the performance consequences of (mis)aligned enforcement. We focus on buyers’ informal (or private) enforcement and not on public enforcement, such as via courts.

To clarify the process by which enforcement takes place, we turn to an example. Frequently, buying firms postpone payments as an enforcement behavior. Bungee Loyalty Programs LLC (http://www.bungeeloyaltyprograms.com) is a US-based firm that provides loyalty programs through the integration of complex software. Bungee Loyalty Programs had agreements with its supplier on the delivery of such software. When substantial problems arose in a software purchase, the firm saw a need to postpone payments as a direct result of perceived problems. Once the payments were postponed, the supplier resolved problems, and Bungee Loyalty Program’s satisfaction with the resolution of the problems was much improved.1

The generalizable insight from this example is of an ordered series of events as depicted in Fig. 1. Specifically, after the deal and contract are decided on, problems may occur in the transaction. Buying firms may then enforce or may refrain from enforcement. As enforcement is a corrective action aimed at remedying problems, some or even complete problem resolution is likely. Enforcement is consequently reflected in the buyers’ satisfaction with problem resolution.

Governance theory provides several perspectives on enforcement. We define enforcement as a corrective action aimed at remedying problems occurring in the transaction. Various types of enforcement exist. These typically follow a pattern where less severe actions precede more severe actions (Rooks & Snijders, 2001). Less severe actions include seeking resolution of problems by referring to or renegotiating the original agreement (Hart & Moore, 1988). If these actions do not result in acceptable outcomes, more severe actions may be administered in the form of delaying payments (Zbaracki, Ritson, Levy, Dutta, & Bergen, 2004). The type of enforcement studied in this paper is important because it is severe, yet more common than legal action, such as seeking sanctions, mediation, or arbitration.2

2.1. How do contracts impact enforcement?

We believe that different components of the same contract can increase and decrease the likelihood of enforcement as contractual components serve different functions (Anderson & Dekker, 2005). Past work has suggested that multiple components are present in contracts. For example, Argyres and Mayer (2007) suggest that specific components are written into contracts to protect and delineate relations, such as communication, roles, and responsibilities, while other terms protect the specific transaction. Related work in the contracting literature suggests the existence of contract components designed to safeguard the specific relationship, as well as to define the terms of the transaction (Anderson & Dekker, 2005; Chen & Bharadwaj, 2009).

As such, we expect contracts to have multiple components. Relational safeguards are components written into a contract that are designed to protect the parties’ interests in maintaining the relationship with one another for an extended period of time. Such components include intellectual property rights, joint management during the relationship, and how provisions in the contract are updated. Because these components are negotiated into the contract to protect the relationship we describe these components as relational safeguards (Poppo & Zenger, 2002). Such relational safeguards are meaningful in determining future behaviors (Ring & Van de Ven, 1992). Transactional safeguards are designed to protect the specific transaction by countering undesired or opportunistic behaviors (Carson, Madhok, & Wu, 2006). Typical transactional safeguards include sanctions on late payment, supplier liability, and arbitration clauses. Service and warranty safeguards outline service and warranty terms, thus protecting the buyer from faulty service provision. Finally, product and price safeguards concern determination of the technical specifications and prices or changes in price levels, thereby allowing the buyer to be confident in associated costs of the transaction. The argument for this structure of four components is rooted in the control system design literature (Jensen & Meckling, 1992). These four components map well with the control framework. Our first component, relational safeguards, relates closely to the “decision rights and responsibilities” component, which considers maintenance of the relationship. Our second component, transactional safeguards, maps onto Jensen and Meckling’s notion of “rewards and punishments” for maintaining or breaking the transaction. Our other two components, service and warranties, and product and price, address the “performance measures” of the control framework.

The incomplete contracting approach suggests that buying firms emphasize drafting the contract to protect various elements of the

---

1 Interview with the CEO of Bungee Loyalty Programs LLC, August 7th, 2012.

2 Of the 971 transactions included in our dataset, only two transactions resulted in a court case. Prior work suggests that such court cases are rare, likely because courts are not reliable enforcers for various reasons, including equivocality in wording and other uncertainties (Crocker & Masten, 1991). Moreover, taking legal action involves costs, such as legal expenses.
transaction (e.g., terms of delivery, payment terms, product support, and warranties; see Anderson & Dekker, 2005; Kern & Willcocks, 2000) by clarifying what is and what is not allowed. Such clarity easily identifies violations, and prior work implicitly assumes that any violations covered in the contract result in application of enforcement (Bergen et al., 1998). Thus, the presence of contractual components designed to protect specific elements of the transaction should result in a greater likelihood that enforcement is applied. Moreover, when such clauses are included, firms may trigger enforcement to protect their reputations (Carson et al., 2006). The components of the contract that specify details of the transaction, service and warranties, or product and price provide a minimum baseline for performance (Anderson & Dekker, 2005). Such baselines help judge performance by identifying when contract execution falls short of specification, which could trigger enforcement. Therefore, we expect that:

H1a. Contractual components designed to protect specific transaction details, such as transactional safeguards, service and warranties, and product and price, increase the likelihood of enforcement.

Prior literature attributes different roles to contracts protecting specific transaction details and contracts codifying relationships (e.g., Lusch & Brown, 1996; Ring & Van de Ven, 1992). We expect that contractual components designed to protect specific elements of the relationship decrease the likelihood of enforcement. Emphasizing relational safeguards that protect the relationship allows issues to be resolved informally since guidelines on how to interact are in place (Ring & Van de Ven, 1992). Thus, we contend that terms designed to insulate the actors from conflict (via relational safeguards) will result in less need to enforce the transaction. Moreover, emphasizing relational elements likely results in a “shadow of the future” between transacting parties because both parties have to consider how they codify future cooperation (Heide & Miner, 1992). Such a shadow of the future helps cooperation and a prudent buyer may, therefore, wish to exercise restraint in enforcement to avoid damaging such future expectations. Thus, we hypothesize that:

H1b. Contractual components designed to protect specific elements of the relationship, such as relational safeguards, decrease the likelihood of enforcement.

2.2. What are the performance consequences of aligned enforcement?

Governance includes elements of establishing and structuring exchange relationships, as well as aspects of enforcement (Heide, 1994). Governance and resultant outcomes are often explained based on a discriminating alignment view. The key aspect of discriminating alignment is that the relation between governance and performance cannot be accurately assessed without inclusion of the factors that lead transacting parties to adopt one form of organization over another (Masten, 1993). Alignment occurs if the factors that governance theories suggest parties to make governance choices are also predictive of such choices. For example, in a make-or-buy context, factors such as asset specificity and uncertainty are argued to predict make rather than buy. Parties heeding such predictions should have superior performance. Thus governance choices are not superior per se (i.e., make is always better than buy). It is the alignment between transaction attributes and governance choice that results in superior performance. This selectivity in choice is discriminating alignment. This discriminating alignment perspective is not specific to TCE as is evident from recent work that adopts broader governance perspectives on alignment (e.g., Bercovitz, Jap, & Nickerson, 2006).

Empirically, the critical issue when testing for discriminating alignment is that predicted governance choices lead to better performance than non-predicted choices. Such analysis requires comparing performance for the predicted enforcement choice (enforce or not) against the actual choice. This results in four scenarios (see Table 1 for an overview of comparison points): a buying firm may make an enforcement (or no enforcement) choice that is aligned to circumstances (Situations 1a and 2a in Table 1). A buying firm may also make a misaligned choice (Situations 1b and 2b in Table 1).

If alignment occurs (i.e., the model-predicted choice is the actual choice), discriminating alignment suggests more satisfactory outcomes than when no alignment occurs (Sampson, 2004). No enforcement, when aligned or predicted, could help the buyer and seller informally resolve problems, thereby enhancing satisfaction. Enforcement not called for (misaligned) could lead to literal interpretation, opportunism, conflicts, or retaliation (Wathne & Heide, 2000; Williamson, 1991). Another reason why alignment should help satisfaction is that behaving according to governance predictions should provide legitimacy. Acting in accordance with what other firms do enhances legitimacy as these behaviors are accepted (Deephouse, 1996). In the face of this conferred legitimacy, the buyer is less likely to experience retaliation due to the buyers’ enforcement, thus helping satisfaction. Past work suggests that aligned governance minimizes the joint costs of governance (such as formalizing the relationship too much) and opportunism, thereby promoting collaborative benefits (Sampson, 2004). In turn, such benefits should enhance satisfaction, compared to situations where these benefits are not realized. In line with this work, meta-analytical evidence suggests that governance choices, when properly aligned with transaction attributes, enhance outcomes including satisfaction (Geyskens et al., 2006). Thus, we posit that:

H2. Higher satisfaction with problem resolution occurs when enforcement is aligned (compared to when enforcement is misaligned).

2.3. What are the performance consequences of misaligned enforcement?

Hypothesis H2 argues that aligned governance results in better outcomes than misaligned governance. This hypothesis compares absolute levels of satisfaction with problem resolution. There are two
cases of misalignment: misaligned enforcement and misaligned non-enforcement (Situations 1b and 2b in Table 1, respectively). Taking a perspective relative to the aligned choice, we argue that the drop in satisfaction with problem resolution is greater for misaligned enforcement. We argue that a situation where enforcement is not expected, yet chosen, results in a greater relative loss of satisfaction than situations where enforcement is expected, yet not chosen. This corresponds to comparisons of Situation 1a against 1b and 2a against 2b in Table 1. Specifically, we argue that the difference in satisfaction with problem resolution between 1a and 1b is greater than between 2a and 2b.

There are arguments why this is expected. First, in a situation where the circumstances (e.g., transaction problems, contracts in place, and complexity of the deal) suggest that enforcement should not be used, unexpectedly enforcing may foster reactance (Brown, Lusch, & Nicholson, 1995). Such reactance is unlikely to be helpful in enhancing buyer satisfaction with the deal. In fact, past work suggests that such reactance may create opportunism as the party punished by enforcement wishes to assert its independence (Brown, Dev, & Lee, 2000). Second, not choosing to enforce may help maintain existing trust between parties. Jeffries and Reed (2000) suggest that trust helps reduce problems through adaptation, thereby increasing satisfaction. The downside risk of choosing enforcement in such a situation is likely much higher compared to situations where the relationship is not functioning well and where enforcement may be expected, yet is not chosen. Thus, we expect that:

H3. The decrease in satisfaction with problem resolution is greater for misaligned enforcement than for misaligned non-enforcement.

3. Method

3.1. Data

We analyze our research questions using the External Management of Automation dataset developed by the Department of Sociology of Utrecht University, The Netherlands, and made available by the Steinmetz archive.3 The dataset was collected in 1995 using surveys with the goal of examining how buying firms manage their IT transactions. It contains detailed information on 971 transactions of information technology products and services (Buskens, Raub, & Weesie, 2000). Because the data were collected in 1995, one could question if the findings from such data still apply today; however, it is very likely that they do since our variables and theory relate to concepts applicable today. Empirical evidence also suggests that IT problems are still pervasive. Approximately 70% of current IT transactions do not deliver on their promises (Laudon & Laudon, 2010); a figure that is very close to the number of transactions reporting (74.06%) problems in our data.

The sampling frame was obtained from the Cendris/Directview database, which contains annually updated information on 100,000 companies, spanning 80% of all Dutch small and medium firms. The study used a 3 × 3 × 4 stratified sampling design, based on three levels of embeddedness (low, medium, high), buyer IT expertise (low, medium, and high), and four IT product categories (standard hardware, standard software, complex hardware, and complex software) to select firms. The study includes at least 15 randomly selected firms per cell. Further, the study deliberately oversampled complex hardware and software transactions, which increases the precision of the estimates for those types of transactions.4

From the sampling frame, 1798 purchasing companies were contacted by phone to identify a sufficiently knowledgeable informant who could provide information on the IT transactions conducted by the firm during the past five years. From these transactions, one random transaction per firm was selected that met the aforementioned stratification criteria and that was completed and independently purchased. This resulted in 1325 usable firms that were the subsequent focus of the study.

Potential participants were subsequently contacted and asked to respond to a survey on the randomly chosen transaction. There were 788 participants, resulting in a favorable net response of 59%. At 547 of the companies, a qualified interviewer administered the survey on-site while 241 companies self-administered the questionnaire after receiving it in the mail. In 183 cases, the informants were willing to answer a second survey on a separate (but also randomly selected) transaction. This resulted in 971 total observations.

The data collected appear to originate from highly competent informants; about 95% of all informants had, on average, 10 years of tenure at their company. The buying firms belong to a wide spectrum of industries, such as logistics, manufacturing, and wholesaling. Over 95% of the responses originated from firms with 66 to 100 employees.

A formal study was conducted to assess non-response by comparing key characteristics of respondent and non-respondent firms, such as location, size, industry, and satisfaction with the transactions of respondents and non-respondents. No significant differences were found, suggesting that non-response issues are of little concern.5 The different methods of data collection (on-site versus mail) were compared and no systematic patterns of differences were observed. Finally, the original stratification criteria were matched with the post hoc data to check the original design goals. The results indicate that the stratification goals were reached.

3.2. Measurement

We now discuss how our variables were measured. Table 2 includes measurement items and scales. Tests for common method bias were executed prior to using this dataset and were found satisfactory. Moreover, reliability and convergent validity were assessed and were also found to be satisfactory (cf. Mooi & Ghosh, 2010).

3.2.1. Independent variables

Enforcement is a corrective action aimed at remedying problems (Antia et al., 2006). Consistent with this theoretical definition, we operationalize enforcement as the buyer’s intentional withholding of payments in response to the buyer perceptions of problems occurring in the product. It is measured using a binary item asking whether the buyer postponed payments as a response to problems occurring in the product. Descriptive statistics show that enforcement is relatively prevalent, occurring in 20.49% of the transactions included in the sample. It is important to note that the participant responded to the enforcement item only if transaction problems were reported (see the description of Transaction problems for further details).

Satisfaction with problem resolution is measured using a single item asking informants how satisfied they were with the way in which problems with this product were solved. This question was also put to the participant only if problems were reported, and followed the enforcement question. Satisfaction, a frequently used measure of inter-organizational performance (Geyskens & Steenkamp, 1999), is a useful measure to study alignment because it provides an evaluation of performance. Satisfaction is useful because the range of frictions created due to transaction costs should be lower when alignment occurs, thus resulting in higher satisfaction as compared to when alignment is absent.

3.2.2. Dependent variables

In addition to these two endogenous variables, our analysis includes a set of antecedent and control variables explaining enforcement and/or satisfaction with problem resolution.

---

4 The sampling process is described in further detail by Buskens and Batenburg (2000).
5 Issues of (non)response are described in further detail by Buskens and Batenburg (2000).
**Table 2**
Measurement.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction problems</td>
<td>Listed below are some potential problems that could arise regarding the product, its delivery, and its service. Please indicate how severe the problems were for each of these items.</td>
</tr>
<tr>
<td>5-point scale (1 = very little;</td>
<td>1. went over delivery schedule</td>
</tr>
<tr>
<td>5 = very much)</td>
<td>2. went over price/budget</td>
</tr>
<tr>
<td></td>
<td>3. product incomplete</td>
</tr>
<tr>
<td></td>
<td>4. product too slow/limited</td>
</tr>
<tr>
<td></td>
<td>5. deviations from specifications made</td>
</tr>
<tr>
<td></td>
<td>6. incompatibility with other IT products</td>
</tr>
<tr>
<td></td>
<td>7. installation too hurried/sloppy</td>
</tr>
<tr>
<td></td>
<td>8. support too slow/late</td>
</tr>
<tr>
<td></td>
<td>9. service too slow/late</td>
</tr>
<tr>
<td></td>
<td>10. necessary adjustments and customization too slow/late</td>
</tr>
<tr>
<td></td>
<td>11. incomplete/unclear documentation</td>
</tr>
<tr>
<td>Enforcement</td>
<td>Did your firm postpone payments as a response to problems regarding this product?</td>
</tr>
<tr>
<td>Binary yes/no response format</td>
<td>Generally, have problems regarding this product been resolved to the satisfaction of your company?</td>
</tr>
<tr>
<td>Satisfaction with problem</td>
<td>See Table 3.</td>
</tr>
<tr>
<td>resolution</td>
<td>What are the different kinds of products/services procured under this agreement? (Multiple answers possible):</td>
</tr>
<tr>
<td>5-point scale (1 = almost never;</td>
<td>1. standard software (1)</td>
</tr>
<tr>
<td>5 = almost always)</td>
<td>2. personal computers (1)</td>
</tr>
<tr>
<td></td>
<td>3. work-stations (1)</td>
</tr>
<tr>
<td></td>
<td>4. peripherals (1)</td>
</tr>
<tr>
<td></td>
<td>5. cabling (1)</td>
</tr>
<tr>
<td></td>
<td>6. network configuration (2)</td>
</tr>
<tr>
<td></td>
<td>7. mini-computer (2)</td>
</tr>
<tr>
<td></td>
<td>8. mainframe (2)</td>
</tr>
<tr>
<td></td>
<td>9. computer driven machines (2)</td>
</tr>
<tr>
<td></td>
<td>10. industry specific software (3)</td>
</tr>
<tr>
<td></td>
<td>11. education (3)</td>
</tr>
<tr>
<td></td>
<td>12. instruction/training (3)</td>
</tr>
<tr>
<td></td>
<td>13. documentation (3)</td>
</tr>
<tr>
<td></td>
<td>14. support (4)</td>
</tr>
<tr>
<td></td>
<td>15. specialized software (4)</td>
</tr>
<tr>
<td></td>
<td>16. consulting (4)</td>
</tr>
<tr>
<td></td>
<td>17. design (5)</td>
</tr>
<tr>
<td></td>
<td>18. customized software (5)</td>
</tr>
<tr>
<td></td>
<td>Does the transaction include only software or hardware components (taken from the</td>
</tr>
<tr>
<td></td>
<td>transaction description)?</td>
</tr>
<tr>
<td></td>
<td>2. How difficult was it to compare this product/service to similar products?</td>
</tr>
<tr>
<td></td>
<td>3. How difficult was it to compare the price/quality ratio of potential suppliers’ products/services?</td>
</tr>
<tr>
<td></td>
<td>If the product failed and had to be replaced, what would be the loss, in terms of time and money, associated with:</td>
</tr>
<tr>
<td></td>
<td>1. purchasing a new product;</td>
</tr>
<tr>
<td></td>
<td>2. training your personnel;</td>
</tr>
<tr>
<td></td>
<td>3. data and information entry;</td>
</tr>
<tr>
<td></td>
<td>4. stoppage at production departments.</td>
</tr>
<tr>
<td></td>
<td>18. customized software (5)</td>
</tr>
<tr>
<td></td>
<td>How important was this product for:</td>
</tr>
<tr>
<td></td>
<td>1. The IT in your company?</td>
</tr>
<tr>
<td></td>
<td>2. The profitability of your company.</td>
</tr>
<tr>
<td></td>
<td>1. Number of potential suppliers at time of purchase</td>
</tr>
<tr>
<td></td>
<td>2. Number of alternative products at time of purchase</td>
</tr>
<tr>
<td></td>
<td>How great was the estimated dependence of your firm on the supplier before the purchase?</td>
</tr>
<tr>
<td></td>
<td>How many years did your company do business with this supplier prior to this transaction?</td>
</tr>
<tr>
<td></td>
<td>Number of employees working for buyer at time of purchase</td>
</tr>
<tr>
<td></td>
<td>Number of employees working for supplier at time of purchase</td>
</tr>
<tr>
<td>Software and hardware</td>
<td>Listed below are some potential problems that could arise regarding the product, its delivery, and its service. Please indicate how severe the problems were for each of these items.</td>
</tr>
<tr>
<td>Measurement ambiguity</td>
<td>1. went over delivery schedule</td>
</tr>
<tr>
<td>Cronbach’s α = .80</td>
<td>2. went over price/budget</td>
</tr>
<tr>
<td>5-point scale (1 = very easy;</td>
<td>3. product incomplete</td>
</tr>
<tr>
<td>5 = very difficult)</td>
<td>4. product too slow/limited</td>
</tr>
<tr>
<td></td>
<td>5. deviations from specifications made</td>
</tr>
<tr>
<td></td>
<td>6. incompatibility with other IT products</td>
</tr>
<tr>
<td></td>
<td>7. installation too hurried/sloppy</td>
</tr>
<tr>
<td></td>
<td>8. support too slow/late</td>
</tr>
<tr>
<td></td>
<td>9. service too slow/late</td>
</tr>
<tr>
<td></td>
<td>10. necessary adjustments and customization too slow/late</td>
</tr>
<tr>
<td></td>
<td>11. incomplete/unclear documentation</td>
</tr>
<tr>
<td>Buyer lock-in</td>
<td>Did your firm postpone payments as a response to problems regarding this product?</td>
</tr>
<tr>
<td>Cronbach’s α = .76</td>
<td>Generally, have problems regarding this product been resolved to the satisfaction of your company?</td>
</tr>
<tr>
<td>5-point scale (1 = very small;</td>
<td>See Table 3.</td>
</tr>
<tr>
<td>5 = very large)</td>
<td>What are the different kinds of products/services procured under this agreement? (Multiple answers possible):</td>
</tr>
<tr>
<td></td>
<td>1. standard software (1)</td>
</tr>
<tr>
<td></td>
<td>2. personal computers (1)</td>
</tr>
<tr>
<td></td>
<td>3. work-stations (1)</td>
</tr>
<tr>
<td></td>
<td>4. peripherals (1)</td>
</tr>
<tr>
<td></td>
<td>5. cabling (1)</td>
</tr>
<tr>
<td></td>
<td>6. network configuration (2)</td>
</tr>
<tr>
<td></td>
<td>7. mini-computer (2)</td>
</tr>
<tr>
<td></td>
<td>8. mainframe (2)</td>
</tr>
<tr>
<td></td>
<td>9. computer driven machines (2)</td>
</tr>
<tr>
<td></td>
<td>10. industry specific software (3)</td>
</tr>
<tr>
<td></td>
<td>11. education (3)</td>
</tr>
<tr>
<td></td>
<td>12. instruction/training (3)</td>
</tr>
<tr>
<td></td>
<td>13. documentation (3)</td>
</tr>
<tr>
<td></td>
<td>14. support (4)</td>
</tr>
<tr>
<td></td>
<td>15. specialized software (4)</td>
</tr>
<tr>
<td></td>
<td>16. consulting (4)</td>
</tr>
<tr>
<td></td>
<td>17. design (5)</td>
</tr>
<tr>
<td></td>
<td>18. customized software (5)</td>
</tr>
<tr>
<td></td>
<td>Does the transaction include only software or hardware components (taken from the</td>
</tr>
<tr>
<td></td>
<td>transaction description)?</td>
</tr>
<tr>
<td></td>
<td>2. How difficult was it to compare this product/service to similar products?</td>
</tr>
<tr>
<td></td>
<td>3. How difficult was it to compare the price/quality ratio of potential suppliers’ products/services?</td>
</tr>
<tr>
<td></td>
<td>If the product failed and had to be replaced, what would be the loss, in terms of time and money, associated with:</td>
</tr>
<tr>
<td></td>
<td>1. purchasing a new product;</td>
</tr>
<tr>
<td></td>
<td>2. training your personnel;</td>
</tr>
<tr>
<td></td>
<td>3. data and information entry;</td>
</tr>
<tr>
<td></td>
<td>4. stoppage at production departments.</td>
</tr>
<tr>
<td></td>
<td>18. customized software (5)</td>
</tr>
<tr>
<td></td>
<td>How important was this product for:</td>
</tr>
<tr>
<td></td>
<td>1. The IT in your company?</td>
</tr>
<tr>
<td></td>
<td>2. The profitability of your company.</td>
</tr>
<tr>
<td></td>
<td>1. Number of potential suppliers at time of purchase</td>
</tr>
<tr>
<td></td>
<td>2. Number of alternative products at time of purchase</td>
</tr>
<tr>
<td></td>
<td>How great was the estimated dependence of your firm on the supplier before the purchase?</td>
</tr>
<tr>
<td></td>
<td>How many years did your company do business with this supplier prior to this transaction?</td>
</tr>
<tr>
<td></td>
<td>Number of employees working for buyer at time of purchase</td>
</tr>
<tr>
<td></td>
<td>Number of employees working for supplier at time of purchase</td>
</tr>
</tbody>
</table>

**Contract terms:** the dataset contains a 24-item scale that measures whether certain contractual terms were present (or not) in the original written contract. These 24 terms, generated by IT managers and lawyers who specialize in contracting, reflect typical contracting practices (see Rooks, Raub, & Tazelaar, 2006). Two senior legal scholars who specialize in business contracting verified this. These 24 terms pertain to the financial, legal, and operational elements used to specify terms of trade and management of the transaction. All individual terms are included in Table 3. The three most commonly included terms relate to price level, payment terms, and warranties. Typically, studies on contracts focus on the overall level of contract completeness or on individual terms. In this paper, we use the same dataset as Anderson and Dekker (2005) and, similar to their approach, we conduct an analysis of the structure of the contract to understand terms commonly used in combination. Using components has distinct benefits. A single summed scale ignores
the complexity and interrelatedness of the terms present in the contract. On the other hand using 24 separate terms ignores that individual terms are commonly used together to reduce transaction hazards, create understanding, and delineate the deal.

Transaction problems are measured as follows: the informant marked if, and to what degree, 11 possible problems occurred in a specific transaction. Using a 5-point scale, the 11 possible problems are weighted by the respondent for their severity. The possible problems include exceeding the quoted price, product and service issues, and implementation issues. The three most common problems were incomplete or unclear documentation, slow or late adjustments and adaptations for implementation, and inadequate support throughout the purchasing, installation, and training process. This variable is a measure of nonperformance and is weighted for the severity of the problems occurring in the transaction on a five-point scale ranging from very little to very much. Of the sample, 241 transactions (25.94%) report no problems being present. Because of the aforementioned routing (questions on enforcement and satisfaction with problem resolution were only put to the respondent if transaction problems were reported), 74.06% of transactions could report on enforcement or satisfaction were only put to the respondent if transaction problems were reported.

Transaction complexity is measured with two indicators. The first is a simple count of the number of products and services covered by the transaction out of 18 types of hardware, software, and services. The second indicator is a scale consisting of five categories that represent products and services that require greater coordination and interaction between the buyer and seller. The scaling is included in brackets behind each item in Table 2. The measurement of transaction complexity is identical to that of Anderson and Dekker (2005). The Software and Hardware item indicates whether the transaction included only software or hardware components. This measure is identical to Vanneste and Puranam (2010). Measurement ambiguity represents the degree of difficulty in defining ex ante and verifying ex post the products and services for which the parties are contracting. We use a three-item scale that taps into the difficulty faced by the buyer in judging the quality of the product/service at the time of delivery, in comparing the focal product/service with other products, and in judging the price/quality ratio of potential suppliers’ products/services (Mooi & Ghosh, 2010). Buyer lock-in is the difficulty faced by the buyer in switching or replacing products or suppliers (Dutta, Bergen, Heide, & John, 1995). We measure it using a four-item scale that looks at the magnitude of costs, in time and money, the buyer would incur if the focal supplier’s product were to be replaced. These damages and costs relate to purchasing another product, (re)training the buyer’s personnel, new data and information entry, and idle production. Transaction importance reflects the value of the transacted products or services to the buyer. Two items measure the buyer’s perceptions of the importance of the products/services to the automation and profitability of the buying organization (Mooi & Ghosh, 2010). Competition is measured using two items focusing on the number of potential suppliers and the number of alternative products at time of purchase. Buyer dependence is a single item measuring the buyer’s perception of dependence on the supplier prior to commencing the transaction. Relationship length accounts for the possible development of relational elements between the buyer-supplier, learning effects, and to generally account for temporal effects (Mayer & Argyres, 2004; Vanneste & Puranam, 2010). It is measured as the numbers of years the buyer and supplier have done business with each other. Buyer and supplier size indicate the number of employees working for the buyer and supplier, respectively.

Table 4 includes correlations, means, and standard deviations for the variables used in this study. These descriptive statistics are calculated for the sample for which there are no missing observations jointly (n = 497).

4. Empirical approach

4.1. How do contracts impact enforcement?

We empirically answer how contracts impact enforcement using a confirmatory analysis of the contract structure, followed by a Probit model to estimate the effects of contractual components on our binary enforcement variable.

We conduct a confirmatory analysis in the form of principal component analysis (PCA) to assess the fit with the component structure expected (Anderson & Dekker, 2005). Specifically, we use PCA to reduce the complexity of the items measuring the various contract

Table 3

<table>
<thead>
<tr>
<th>Terms/components</th>
<th>Relational safeguards</th>
<th>Transactional safeguards</th>
<th>Service and warranties</th>
<th>Product and price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary yes/no response format</td>
<td>.54</td>
<td>.86</td>
<td>.82</td>
<td>.73</td>
</tr>
<tr>
<td>Intellectual property (50%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piracy protection (29%)</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrictions on product use (32%)</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-disclosure (14%)</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reservation of spare-parts (18%)</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Updating (45%)</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculation of R&amp;D costs (19%)</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint management during transaction (22%)</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanctions on late payment (26%)</td>
<td></td>
<td>.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liability supplier (55%)</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Force majeure supplier (32%)</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurance supplier (24%)</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arbitration (27%)</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms of notice (32%)</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warranties supplier (80%)</td>
<td></td>
<td></td>
<td>.81</td>
<td>.82</td>
</tr>
<tr>
<td>Quality (norms) (38%)</td>
<td></td>
<td>.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration service (74%)</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration maintenance (64%)</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery time (70%)</td>
<td></td>
<td>.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price determination (79%)</td>
<td></td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price level (88%)</td>
<td></td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price changes (40%)</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical specifications (70%)</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payment terms (80%)</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KR-20</td>
<td>.80</td>
<td>.82</td>
<td>.73</td>
<td>.70</td>
</tr>
</tbody>
</table>

The frequency of occurrence of each term is listed in brackets behind the name of each term. N = 971.
Correlations > |.10| are significant for product performance and product maintenance. Unforeseen circumstances, and liabilities. Protect it and the specification is higher, buyers more likely choose enforcement because important transactions typically affect importance of summing all contract terms. Our results indicate that the 24 contract terms represent four interpretable components, thereby rejecting the idea that our contract items represent a single dimension. These findings are consistent with those of Anderson and Dekker (2005).7

We describe the four components. Relational safeguards protect the parties’ abilities to work with one another. These terms define the constraints placed on the interaction of the buyer and seller and include items related to intellectual property and the continued use of the product. Transaction safeguards are mostly focused on the supplier and protect it and the specific transaction from, for instance, cancelation, unforeseen circumstances, and liabilities. Service and warranties establish expectations for product performance and product maintenance. Finally, product and price define the technical specifications, the compensation made in return, and how potential changes in compensation are dealt with.

Nomological validity is established by inspecting the correlations among these four contractual components and known transaction attributes, such as transaction complexity and measurement ambiguity. TCE predicts that, as transaction complexity and measurement ambiguity increase, contracts become more complete (Geyskens et al., 2006; Rindfleisch & Heide, 1997). Table 4 shows these correlations and, as expected, positive and significant correlations are present, supporting nomological validity.

We also calculated the Kuder–Richardson Formula 20 (KR-20) for each component. The KR-20 is a measure of internal consistency for binary items. The KR-20 coefficients in Table 3 (ranging from .70 to .82) indicate acceptable consistency.

We use the four contractual components and a set of control variables to explain enforcement using a Probit model. Past work based on the same dataset (cf. Anderson & Dekker, 2005; Mooi & Ghosh, 2010; Vanneste & Puranam, 2010) provides an extensive set of variables to explain governance choices. Relying on previously used predictors should reduce omitted variable bias or other model misspecification issues. We discuss these in turn. Transactional problems are a key reason why buyers choose to enforce transactions. Enforcement may be chosen because the buyer expects it will remedy problems that might otherwise not be resolved. Transaction complexity increases the need for coordination between the buyer and seller. Through enforcement, buyers may signal to sellers that coordination efforts need to be undertaken. In the context of IT, transactions that consist of only software or hardware components might be easier to coordinate than transactions that require a combination. Measurement ambiguity causes a performance evaluation problem for buyers that could cause opportunistic behavior on the part of the seller, which buyers might counter through enforcement. An alternative view is that measurement ambiguity makes it more difficult to assign cause and effect or blame, thus reducing the ability of the buyer to enforce the transaction. Buyer lock-in implies a greater difficulty switching to an alternative supplier or product, thereby making enforcement aimed at a particular supplier more complicated. When importance of the transaction is higher, buyers more likely choose enforcement because important transactions typically affect important outcomes, such as profitability. If greater competition is present, enforcement may be easier because alternatives are available for future transactions. Buyer dependence on the supplier could reduce the likelihood of buyer-imposed enforcement due to the risks of supplier retaliation. A longer relationship length may have fostered elements of experience, trust, or relational exchange that help buyers resolve issues based on dialog (Sheng, Brown, Nicholson, & Poppo, 2006). Buyer and supplier size are proxies for the power of the two respective sides

Table 4
Correlations, means, and standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Enforcement</td>
<td>.94</td>
<td>.46</td>
<td>.13</td>
<td>.10</td>
<td>.08</td>
<td>.16</td>
<td>.07</td>
<td>.21</td>
<td>.00</td>
<td>.03</td>
<td>.16</td>
<td>.22</td>
<td>.10</td>
<td>.04</td>
<td>.00</td>
<td>.13</td>
<td>.25</td>
<td>.50</td>
</tr>
<tr>
<td>2. Satisfaction with problem resolutionb</td>
<td>.03</td>
<td>.00</td>
<td>.10</td>
<td>.11</td>
<td>.05</td>
<td>.07</td>
<td>.03</td>
<td>.07</td>
<td>.08</td>
<td>.03</td>
<td>.06</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>.06</td>
<td>.08</td>
</tr>
<tr>
<td>3. Transaction problemsb</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>4. Relational safeguards</td>
<td>.07</td>
<td>.13</td>
<td>.10</td>
<td>.03</td>
<td>.05</td>
<td>.06</td>
<td>.02</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td>5. Transaction safeguards</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>6. Service and warranties</td>
<td>.08</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>7. Product and price</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>8. Transaction complexity</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>9. Software</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>10. Hardware</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>11. Measurement ambiguity</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>12. Buyer lock-in</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>13. Transaction importance</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>14. Competition</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>15. Buyer dependence</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>16. Relationship length</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>17. Buyer size</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>18. Supplier size</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Correlations > |.10| are significant at p < .05, two-tailed.

n = 497.

a The correlations in these columns are with binary variables and, therefore, indicative only.
b These variables were reported on only if at least one transaction problem occurred, which is the case for 74.06% of our observations.

7 The structure and assignment of items is identical when factor analysis is used.
to the transaction. We include all of these controls in our model and estimate our Probit model as follows:

\[
    \text{Prob}(Y_i = 1) = \Phi(X_i \beta).
\]

\[Y_i = 1\] if the seller delays payments and \( \Phi \) represents the cumulative normal distribution. \( Y_i \) is explained by a vector of aforementioned explanatory variables, \( X_i \), and their weights, \( \beta \). As our dataset includes 183 buyers that report on two transactions, we use a cluster-robust estimator for our Probit model to account for intra-firm correlations (Wooldridge, 2002). We use this estimator because unobserved firm-specific factors, such as corporate standards and management experience, may cause correlated enforcement practices within firms. The results of the estimation process are included in Table 5.

The Probit model resulting from the aforementioned estimation setup is highly significant. The parameter estimates suggest that transaction safeguards (\( p < .05 \)) and service and warranties (\( p < .05 \)) increase the likelihood of enforcement, consistent with H1a. We find no effect of product and price on enforcement. Relational safeguards reduce the likelihood of enforcement (\( p < .05 \)), consistent with H1b.

We also estimate all possible interaction effects between our contracting components, as well as interactions between our contracting components and transaction characteristics, but no pattern of significant interactions (\( p > .05 \)) appears.

4.2. What are the performance consequences of aligned enforcement?

We investigate if buying firms correctly enforce by answering what the outcomes of enforcement are in the form of satisfaction with problem resolution, and to what degree selective enforcement matters.

Our context and approach for testing discriminating alignment raises two issues that make OLS a poor analysis tool. The first issue is that, based on TCE considerations, we expect buyers to make an enforcement choice based on resultant outcomes. This means that enforcement should be endogenized, rather than treated as an exogenous variable. The second issue is that we only observe outcomes for choices made but not for choices not made. That is, we compare outcomes for different buyers for different transactions, instead of comparing outcomes for the same buyer, for the same transaction. Thus, in econometric terms we face issues of endogeneity in explaining enforcement. Endogeneity yields biases whose size and direction are difficult to predict (see Chapter 9 of Maddala, 1983 for a proof) and potentially leads to faulty conclusions.

Heckman (1979) addresses this problem using a single Probit model (the choice equation, as introduced previously) to endogenize the choice, followed by a single OLS regression model that includes a correction term—the inverse Mills ratio—to deal with potential endogeneity by correcting the outcome. We use a modification of the Heckman approach with two outcome equations because there is no theory to assume a priori that satisfaction is influenced by the same antecedents to the same degree for enforcement versus no enforcement. This modification to the Heckman approach, which uses two outcome models, is the endogenous switching regression model. This model has been applied to situations where binary choices impact performance (see for example Carson et al., 2006; Leiblein, Reuer, & Dalsace, 2002).

Table 5

<table>
<thead>
<tr>
<th>Drivers of enforcement.</th>
<th>Parameter estimates with standard errors in parentheses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>−4.415 (−6.39)*</td>
<td></td>
</tr>
<tr>
<td>Transaction safeguards</td>
<td>.094 (.044)*</td>
<td></td>
</tr>
<tr>
<td>Service and warranties</td>
<td>.120 (.068)*</td>
<td></td>
</tr>
<tr>
<td>Product and price</td>
<td>.013 (.054)</td>
<td></td>
</tr>
<tr>
<td>Control variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relational safeguards</td>
<td>−1.500 (−4.02)*</td>
<td></td>
</tr>
<tr>
<td>Transaction problems</td>
<td>.101 (.009)*</td>
<td></td>
</tr>
<tr>
<td>Transaction complexity</td>
<td>.017 (.022)</td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>.179 (.204)</td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td>−.267 (−2.26)</td>
<td></td>
</tr>
<tr>
<td>Measurement ambiguity</td>
<td>.029 (.030)</td>
<td></td>
</tr>
<tr>
<td>Buyer lock-in</td>
<td>.016 (.021)</td>
<td></td>
</tr>
<tr>
<td>Transaction importance</td>
<td>.029 (.043)</td>
<td></td>
</tr>
<tr>
<td>Competition</td>
<td>.006 (.035)</td>
<td></td>
</tr>
<tr>
<td>Buyer dependence</td>
<td>.054 (.060)</td>
<td></td>
</tr>
<tr>
<td>Relationship length</td>
<td>−.012 (.013)</td>
<td></td>
</tr>
<tr>
<td>Buyer size</td>
<td>−.009 (.070)</td>
<td></td>
</tr>
<tr>
<td>Supplier size</td>
<td>.049 (.052)</td>
<td></td>
</tr>
<tr>
<td>Model fit</td>
<td>Wald ( x^2 = 103.06, p &lt; .000 )</td>
<td></td>
</tr>
<tr>
<td>Estimation method</td>
<td>Probit, with cluster-robust standard errors, ( n = 703 )</td>
<td></td>
</tr>
</tbody>
</table>

All tests are two-tailed.

\[ * p < .05. \]

The correction terms to account for endogeneity are the inverse Mills ratios and are constructed from the predictions of the enforcement model (see Table 5). Following Heckman (1979) we compute the following ratios:

\[
    \lambda_{\text{no enforcement}} = -\phi(\beta'X_i) / (1 - \phi(\beta'X_i))
\]

\[
    \lambda_{\text{enforcement}} = \phi(\beta'X_i) / (\Phi(\beta'X_i))
\]

where \( \phi \) is the standard normal probability function and \( \Phi \) is the standard normal cumulative density function. The two \( \lambda \) terms are entered into the outcome models to control for endogeneity. As we estimate two \( \lambda \) terms, we can separately interpret the strength and direction of endogeneity for buyers who choose to enforce and those who choose not to. Note that we exclude relationship length and buyer/supplier size to identify the model, since there is no clear expectation why these variables should influence satisfaction with a particular product per se. An inspection of the correlations also shows

Table 6

<table>
<thead>
<tr>
<th>Performance implications of enforcement.</th>
<th>Satisfaction with problem resolution</th>
<th>Satisfaction with problem resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>No enforcement condition</td>
<td>Enforcement condition</td>
</tr>
<tr>
<td>Constant</td>
<td>.871 (2.351)</td>
<td>.4377 (1.080)*</td>
</tr>
<tr>
<td>Inverse Mills ratio</td>
<td>−1.097 (−5.39)*</td>
<td>−1.214 (−5.02)*</td>
</tr>
<tr>
<td>Transaction safeguards</td>
<td>.090 (−.49)</td>
<td>.038 (.052)</td>
</tr>
<tr>
<td>Service and warranties</td>
<td>.093 (0.71)</td>
<td>.048 (.064)</td>
</tr>
<tr>
<td>Product and price</td>
<td>.050 (.038)</td>
<td>.066 (.066)</td>
</tr>
<tr>
<td>Relational safeguards</td>
<td>−.086 (.066)</td>
<td>−.069 (.072)</td>
</tr>
<tr>
<td>Transaction problems</td>
<td>.009 (.043)</td>
<td>−.011 (.037)</td>
</tr>
<tr>
<td>Transaction complexity</td>
<td>.050 (.015)*</td>
<td>.006 (.021)</td>
</tr>
<tr>
<td>Software</td>
<td>.204 (.149)</td>
<td>.207 (.189)</td>
</tr>
<tr>
<td>Hardware</td>
<td>−.215 (.182)</td>
<td>−.189 (.320)</td>
</tr>
<tr>
<td>Measurement ambiguity</td>
<td>.009 (.027)</td>
<td>.013 (.034)</td>
</tr>
<tr>
<td>Buyer lock-in</td>
<td>.030 (.021)</td>
<td>.003 (.026)</td>
</tr>
<tr>
<td>Transaction importance</td>
<td>.008 (.029)</td>
<td>.015 (.053)</td>
</tr>
<tr>
<td>Buyer dependence</td>
<td>.059 (.043)</td>
<td>.002 (.068)</td>
</tr>
<tr>
<td>Competition</td>
<td>.017 (.021)</td>
<td>−.011 (.044)</td>
</tr>
<tr>
<td>Model fit</td>
<td>( R^2 = .30 )</td>
<td>( R^2 = .44 )</td>
</tr>
<tr>
<td>Estimation method</td>
<td>Ordinary Least Squares</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td></td>
<td>using cluster-robust standard errors, ( n = 553 )</td>
<td>using cluster-robust standard errors, ( n = 150 )</td>
</tr>
</tbody>
</table>

\[ * p < .05. \]
that these variables do not correlate significantly with satisfaction with problem resolution. Even though endogenous switching regression models are identified without exclusion restrictions through the non-linearity of the Probit model, this tends to result in multicollinearity and inflated standard errors (Bushway, Johnson, & Slocum, 2007). When excluding identifying variables, our results indicate that multicollinearity is of little concern.

Table 6 shows the results of estimating the outcomes of enforcement. In these models, the two Mills ratios are included, as are all contractual components and control variables; relationship length and buyer and supplier size are excluded for the aforementioned identification purposes. Even when controlling for all these variables, the two Mills ratios are significant \((p < .05)\), thereby indicating endogeneity issues are a concern and our approach (accounting for endogeneity) is appropriate. Because both Mills ratios are significant, the unobserved reasons to enforce also influence the outcome. More specifically, because of their negative signs (suggesting higher satisfaction as the inverse Mills ratio is used), buyers have, on average, self-selected into the choice that is most favorable, given the circumstances (Leiblenn et al., 2002; Shaver, 1998). Next to this, transaction complexity \((p < .05)\) increases satisfaction with problem resolution but only for observations with no enforcement.

### 4.3. What are the performance consequences of misaligned enforcement?

Having established that endogeneity is an issue and affects outcomes, we now discuss to what degree selective enforcement matters. What is required to answer this question is to contrast observed choices with the alternatives for the same buyer. The difference between these choices is the foregone satisfaction with problem resolution due to a misaligned choice. Using the endogenous switching regression model, we investigate this foregone satisfaction to shed light on the normative effects of enforcing. That is, when buyers choose to enforce, would their performance have been worse had they chosen not to enforce and vice-versa? We refer to these alternative outcomes as ‘counterfactuals’ (i.e., the outcome a buyer would have achieved had it made the alternative choice). The difference between the observed choice and counterfactuals informs on the costs of misaligned enforcement and are, therefore, of substantive importance. We calculate these counterfactuals following Maddala’s (1983) approach. This approach requires the calculation of four scenarios across two dimensions that conform to Table 1: enforcement (versus no enforcement) and when model predictions are aligned (versus when they are misaligned). Thus, comparisons are made between a buyer’s levels of satisfaction with problem resolution when these are predicted correctly not to enforce transactions versus when they are incorrectly predicted not to enforce. We calculate these four predictions as follows \((\hat{S} \text{ indicating predicted satisfaction with problem resolution; indicators in brackets referring to situations in Table 1})\):

\[
\hat{S}(1a) = \beta \lambda_{\text{no enforcement}} + \beta X
\]

\[
\hat{S}(1b) = \beta \lambda_{\text{no enforcement}} + \beta X
\]

\[
\hat{S}(2a) = \beta \lambda_{\text{enforcement}} + \beta X
\]

\[
\hat{S}(2b) = \beta \lambda_{\text{enforcement}} + \beta X
\]

Regarding the performance consequences of aligned enforcement, buyers who were predicted to enforce and chose to do so (i.e., an aligned situation) experienced an average satisfaction of 3.81. Of the buyers who enforced but were predicted not to enforce, this drops to 3.67. This difference is significant at \(p < .05\) (for details see Table 7). Thus, buyers who were predicted to enforce (aligned), experienced higher satisfaction than if they misaligned by not enforcing, suggesting that aligned enforcement pays off in this condition for the buyer. For the situation where buyers did not enforce, conform to predictions (aligned), the average level of predicted satisfaction with problem resolution is 4.59. This drops to an average predicted level of 3.82 for buyers who did not enforce as predicted (difference significant at \(p < .05\)—for details see Table 7). These findings support H2, since the differences between the aligned and misaligned choice are both significant \((p < .05)\) for the enforcement and no enforcement scenarios.

Regarding the performance consequences of misaligned enforcement, our data suggest that there are substantial costs to making mistakes. Specifically, these costs are relatively highest when enforcement is not expected (per the model), yet chosen. The drop of 0.77 points (on a five-point scale) is significant \((p < .05)\) for the enforcement and no enforcement scenarios. This insight: the use of contracts designed to support relational elements of the transaction itself (such as provisions concerning liability of the supplier), enforcement is more likely. This suggests an interesting hypothesis: the use of contracts designed to support relational elements of the deal reduces the likelihood of enforcement. It is likely that when more relational safeguards are present, recourse is sought in other ways, such as voicing problems to the supplier or by sharing information to remedy problems. When more relational elements are present, these elements could help cooperation and avoid the types of problems that trigger enforcement. Therefore, drafting clauses that safeguard the relationship could be a very effective governance mechanism.

That these components work differently on enforcement suggests that future research on the role of contracts vis-à-vis enforcement should explicitly take into account the different components of contracts. This has interesting implications for the study of contracts. Often contracts have been implicitly assumed to be monolithic by...
5.2. What are the performance consequences of (mis)aligned enforcement?

Regarding our second and third contributions, we build on an endogenous switching regression model to unravel some of the complexities associated with explaining the outcomes of enforcement. Our results show that alignment of enforcement with transactional attributes and contractual components matters because it enhances satisfaction with problem resolution. The two significant inverse Mills ratios clearly suggest that endogeneity is a key concern. Specifically, we find that the unobserved reasons that lead firms to enforce (or not) are those that also help firms avoid lower satisfaction with problem resolution. Thus, firms tend to enforce when this is most beneficial to them. In an enforcement context, we provide evidence that alignment is critical to enhance performance.

The negative and significant correlation between enforcement and satisfaction ($r = -0.28$) might be interpreted as an indication that enforcement reduces satisfaction. Seemingly, this suggests that choosing enforcement is not a rational choice. However, the adopted method accounts for endogeneity and suggests a more subtle and rational explanation: although enforcement may result in poor outcomes, not choosing to enforce when this is expected results in even worse outcomes. Our findings suggest that the adopted method is critical and allows us to obtain new insights not allowed by traditional linear modeling. We also note that calculating the costs of aligned versus misaligned governance through a “what if” analysis that compares a predicted choice against alternatives is helpful in understanding outcomes. Such an analysis allows calculating the “governance effect” of predicted and mistaken (or unpredicted governance) for the same firm and thereby provides normative insight into the value of governance carefully chosen to reflect transaction attributes. Specifically, we find that buying firms that enforce when predicted to do so fare better than those that do not follow such predictions. Buying firms that act conservatively by not enforcing, even if circumstances would likely lead to such an expectation, also appear to not serve their best interests. The signal sent by not enforcing may lead to the seller not remedying problems, since the buyer appears to be able to “get away” by shirking and evading obligations. Suppliers may simply not put in the effort, as there are costs to remediating problems that, in the absence of enforcement, may deter the supplier from remedial actions.

We take our analysis beyond the scope of much of the previous governance work as our adopted method allows us to understand asymmetries in misalignment; that is, is the foregone satisfaction of misaligned enforcement and misaligned non-enforcement the same? Understanding asymmetries is helpful given that decision makers typically weigh losses and gains differently. We find that there are substantial costs to misaligning the choice to enforce. In fact, the relative drop in satisfaction is greatest when buying firms enforce when this is misaligned. Acting aggressively in situations where enforcement is not expected, yet chosen, carries a loss in satisfaction. Likely, such aggressive action engenders much reactance by the seller and sours the relationship. Misaligned enforcement may also create additional problems such as retaliation, mistrust, and erosion of good will.

These asymmetries, both in their absolute and relative effects imply that buying firms that discriminate and enforce fare best in terms of the degree to which transaction problems are satisfactorily resolved.

6. Limitations and further research

One limitation is that our study only considers transaction problems prior to enforcement. Enforcement may create new transaction problems, such as the seller refusing to honor warranties. Future research on the problems that may surface following enforcement would be interesting.

Another limitation is that our satisfaction measure relies on a single survey item. Having multiple items could help validity and allow calculating reliability indicators, such as Cronbach’s $\alpha$. Adopting other outcome measures of enforcement, such as the likelihood of choosing a supplier for repeat business following enforcement, appear interesting research topics.

Finally, our study focuses on a single enforcement action. It is possible that multiple enforcement actions occur. Investigating the reasons for multiple enforcement actions and their outcomes are interesting avenues for further research. Specifically, future research on how private enforcement leads to public enforcement would help the understanding of enforcement substantially as it would help bridge the private enforcement approach often adopted in management and marketing literature with the public enforcement view commonly adopted in the legal literature.

7. Conclusion

A dilemma for buying firms facing transaction problems is whether they should take action to remedy problems. We find that the drafted contracts set the stage for enforcement. Relational safeguards reduce the likelihood of enforcement while transaction safeguards increase it. We also suggest that enforcement, through delaying payment to the supplier, could be helpful to the buyer. However, enforcement is only useful when it aligns with the contract in place and the transaction problems that surface, and is thereby expected. Thus, enforcement is a tool that should be selectively used.

Acknowledgments

The authors are grateful to the department of Sociology, Utrecht University, and the Steinmetz archive for providing access to the data. We also thank the seminar participants at the Wisconsin School of Business, University of Wisconsin-Madison, College of Business, Colorado State University, Aston Business School, Birmingham, and the VU University, Amsterdam, for their useful comments. We also thank Joe Cannon, Shankar Ganesan, and Inge Geykens for their comments. Finally, we thank the AMA for giving an earlier version of this paper the best paper award in interorganizational issues, business-to-business marketing, supply chain management, and channels of distribution in the summer 2008 conference.

References


Customer satisfaction and consumer expenditure in selected European countries

Matthew C.H. Yeung a, Bala Ramasamy b,1, Junsong Chen c,*, Stan Paliwoda d,2

a Lee Shau Kee School of Business and Administration, The Open University of Hong Kong, 30 Good Shepherd Street, Homantin, Kowloon, Hong Kong Special Administrative Region
b China Europe International Business School, 699 Hong Feng Road, Pudong, Shanghai, PR China
c School of Business, East China University of Science and Technology, 130 Meilong Road, Xuhui District, Shanghai 200237, PR China
d Department of Marketing, University of Strathclyde, Stenhouse Building, 173 Cathedral Street, Glasgow G4 0RQ, Scotland, UK

ARTICLE INFO

Article history:
First received in 14, November 2011 and was under review for 8 months
Available online 31 August 2013
Area Editor: Roland T. Rust

Keywords:
Customer satisfaction
Personal consumption expenditure (PCE)
Customer satisfaction index (CSI)
Service economy
Self-expressive values
Economic freedom

ABSTRACT

The relationship between customer satisfaction and company performance has been extensively researched at both the consumer and firm levels. However, little is known about the impact of customer satisfaction at the economy-wide level, especially in Europe. This study aims to link customer satisfaction to personal consumption expenditure using panel data collected from 1999 to 2011 and covering nine European countries. Our findings suggest a significant relationship between customer satisfaction and consumer expenditure in these countries. In addition, economic structure, culture, political economy and socio-economic factors have been examined to understand the impact of cross-country differences on this relationship. The results reflect the importance of satisfied consumers on the economy as a whole; thus, efforts at boosting customer satisfaction should become a national agenda.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The guiding philosophy of marketing management argues that the creation of customer satisfaction is the lifeblood of marketing theory and practice. What people essentially desire is not products but a satisfying experience (Baker et al., 1983). Sales and marketing strategies are therefore centered on creating customer satisfaction (Bond, Fink, & Ross, 2001; Webster, 1992). At the firm level, it has been well documented that customer satisfaction is positively linked to a company’s business performance, profitability and competitive advantage (Anderson & Mittal, 2000; Helgesen, 2006; Oliver, 1997; Yeung & Ennew, 2000). Companies often benefit from engaging in marketing practices that help enhance customer satisfaction (Mittal & Lasser, 1998; Sureshchandar, Rajendran, & Anantharaman, 2002). Therefore, regular customer satisfaction surveys are often undertaken by companies to assess their marketing effectiveness. Measures of customer satisfaction are also often used to determine compensation for executives (Ittnar, Larcker, & Rajan, 1997).

However, it is less clear how an economy as a whole could benefit from improved customer satisfaction. This uncertainty is mainly due to marketing’s inability to quantify its added value to the economy through a longitudinal analysis. Despite some conceptual emphasis on the macroeconomic relevance of customer satisfaction (see, for example, Fornell, Ittner, & Larcker, 1996; Fornell et al., 1996), empirical studies on issues linking customer satisfaction to consumer spending and economic growth are scant. It is only quite recently that researchers have started to elevate customer satisfaction to a macroeconomic variable and analyze its effect on national consumption. One reason for the absence of empirical studies is insufficient customer satisfaction time series data required for analysis. Forrnell, Rust, and Dekimpe (2010) employed an asymmetric growth model to test the effects of changes in the American Customer Satisfaction Index (ACSI) on changes in consumer expenditure. Concurrently, Ramasamy and Yeung (2010) examined the extent to which the ACSI and the Consumer Confidence Index (CCI) can act as determining variables in the national consumption function.

In this study, we extend Fornell et al.’s (2010) and Ramasamy and Yeung’s (2010) work by testing the effects of the Customer Satisfaction Index (CSI) on consumer spending in several European countries. The motivation of this study is not only to affirm previous findings in a different economic setting but also to generate insights into cross-country issues. Specifically, the study intends to address the following research issues: (1) Could CSI act as a significant driver of aggregate consumption expenditure in the selected European countries? (2) Does the influence of customer satisfaction on consumption expenditure vary across different European countries, and what factors account for these differences?

Nine European countries were selected based on consistent data availability. Panel data modeling techniques were employed in this study. Building on Fornell et al.’s (2010) model, we included income,
inflation, debt and consumer confidence as control variables. We also investigated the role of cross-country differences that moderate the significance of the CSI as a driver of consumer spending. Drawing on the theories of several earlier cross-country studies comparing customer satisfaction at the national level (e.g., Morgeson, Mithas, Keiningham, & Aksoy, 2011; Ogikubo, Schvaneveldt, & Enkawa, 2009) as well as studies examining the moderating factors of the satisfaction–purchase link (e.g., Seiders, Voss, Grewal, & Godfrey, 2005), we propose four major moderators: economic structure, culture, social economy, and political economy. The paper attempts to answer questions such as whether the CSI can be a more significant driver for an economy with a more significant service sector, a more educated economy, a higher per capita income, or with greater economic freedom. Pursuing these questions not only is interesting but also provides policy-makers with better insights into the influence of customer satisfaction on the economy. Our findings confirm the role of customer satisfaction as a significant driver of consumer expenditure in the European countries covered in this study. Furthermore, our results suggest that the CSI is a relatively more important driver of consumption in economies where the services sector is more dominant, where the economy is freer, and where education levels are lower. We also find that culture matters; i.e., consumption expenditure is influenced more by the CSI in societies with stronger survival values (as opposed to self-expression values). However, given the data limitations we experienced, our findings have to be considered as preliminary. Nevertheless, our findings, together with those of similar studies, could lead to more in-depth research into the link between macro-marketing and macroeconomic variables.

2. Customer satisfaction at the micro level

Customer satisfaction studies at the micro level explore issues related to the consumer and the firm. At the individual consumer level, studies have focused on the nature, antecedents, moderators and consequences of customer satisfaction, and the variables under consideration are generally individual-level attitudinal or behavioral constructs. Empirical studies have investigated the relationship between customer satisfaction and customer loyalty (Anderson & Sullivan, 1993; Morgan & Rego, 2006; Shankar, Smith, & Rangaswamy, 2003; Taylor & Baker, 1994), customer satisfaction and actual purchase (Bolton, Lemon, & Bramlett, 2006; Seiders et al., 2005; Voss, Godfrey, & Seiders, 2010), and customer satisfaction and share of wallet, which is often treated as another important means of measuring customer loyalty (Baumann, Burton, & Elliott, 2005; Cool, Keiningham, Aksoy, & Hsu, 2007; Keiningham, Perkins-Munn, Aksoy, & Estrin, 2005). Another stream of research investigated the moderating factors of customer satisfaction ratings, including demographic factors (Mittal & Kamakura, 2001), customer experiences with competing firms (Wangenheim & Bayon, 2004), and perceived risk (Johnson, Garbarino, & Sivadas, 2006).

At the firm level, studies have sought to examine the financial value that customer satisfaction brings to the company. For instance, using data from the Swedish Customer Satisfaction Barometer (SCSB) (based on 77 firms representing 70% of Sweden’s economic output), Anderson, Fornell, and Lehmann (1994) found a significant association between customer satisfaction and return on assets (ROA). Yeung and Ennew (2000) linked data from the ACSI to a range of financial performance measures, while Eklof, Hackl, and Westlund (1999), Anderson, Fornell, and Mazvancheryl (2004), and Matzler, Hinterhuber, Daxer, and Huber (2005) empirically demonstrated a relationship between customer satisfaction and shareholder value. Scholars have observed that customer satisfaction boosts shareholder value by increasing cash flow growth and reducing its volatility (Fornell, Mithas, Forrest, & Krishnan, 2006; Grucha & Rego, 2005). In particular, high levels of customer satisfaction should be negatively associated with the cost of debt financing and positively associated with credit rating (Anderson & Mansi, 2009).

In sum, efforts at boosting customer satisfaction increase customer purchase and are financially rewarding to the firm.

3. Customer satisfaction at the macro level

The customer satisfaction literature at the micro level described earlier indicates how consumers respond to satisfactory or unsatisfactory experiences and how companies can benefit from creating satisfactory customer experiences. At the macroeconomic level, it can be argued that customer satisfaction is a driver of aggregate consumer expenditure. From consumers’ perspectives, people tend to spend their incomes in ways that would yield the greatest satisfaction. The utility or satisfaction that consumers derive from previous consumption will therefore affect the expected utility of future purchases (Johnson, Anderson, & Fornell, 1995), increase their consumption expenditure in the following period (Homburg, Koschat, & Hoyer, 2005), and lead to more cross- and up-selling (Li, Sun, & Wilcox, 2005). Furthermore, satisfied consumers will generate positive word-of-mouth, which will increase the confidence of other consumers and encourage more spending (Danaher & Rust, 1996). The practices of monitoring customer satisfaction at the industry level reveal that customer satisfaction is an agenda not only for individual firms but also for the entire industry.

From a macroeconomic perspective, it is argued that while productivity measures the quantity of economic output, customer satisfaction measures the quality of economic output (Fornell, Ittner, et al., 1996). Clearly, consumption will not be sustainable if the quality of output is compromised. More importantly, because aggregate consumption by households contributes a sizeable portion of GDP (for example, 70.1% in the US, 61.8% in the UK and 55.8% in Germany in 2010), increases in consumption expenditure due to increases in satisfaction levels can have a direct impact on the economy as a whole. Therefore, for the policy-maker, customer satisfaction has the “potential to be a useful tool for evaluating and enhancing the health of the nation’s economy, both in terms of national competitiveness and the welfare of its citizens” (Fornell et al., 1996).

Despite this logic, some economists have questioned whether there is indeed a link between overall customer satisfaction and economy-wide consumer-spending trends (Hilsenrath & Freeman, 2002). Their doubt is most likely attributed to the assumption in the economic theory of consumer behavior that regards one unit of satisfaction from a particular consumption package as being independent of the satisfaction derived from other consumption units. More importantly, ever since Keynes introduced personal disposable income as a determinant of income in the General Theory, economists have been pre-occupied with income as the main driver of consumption. From Milton Friedman’s (1957) permanent income hypothesis to Ando and Modigliani’s (1963) Life Cycle Hypothesis, a considerable amount of attention has been devoted to this area. Other determinants of consumer spending have also been considered, including inflation rates, unemployment rates, and stock market performance (see for example, Hendry & von Unger-Sternberg, 1981; Ludvigson & Steindel, 1999; Poterba, 2000). However, these determinants are also indirectly related to future income streams.

Nevertheless, researchers have investigated other drivers of consumer spending aside from income-related variables. For instance, Katona (1975) argued that consumption is a function of disposable personal income plus a variable measuring the aggregate willingness to spend, where such willingness is defined as the relative degree of optimism or pessimism felt by consumers. Consumer sentiment is defined as a measure of willingness based on a perceived future condition. Using the Index of Consumer Sentiment (ICS), Carroll, Fuhrer, and Wilcox (1994) investigated its explanatory power with respect to personal consumption expenditures. Desroches and Gosselin (2002) further argued that the ICS captured information about expected future income and the willingness of consumers to spend based on their perception of future uncertainties. Their theoretical justification was also partly based on Katona’s (1975) extension of the Keynesian consumption function.
The link between consumer sentiment/confidence and consumption expenditure has been empirically tested for several European countries. Belessiotis (1996) for France and Acemoglu and Scott (1994) for the UK suggested that consumer confidence is a leading indicator of future consumption. In the case of the UK, for instance, Delorme, Kammerschen, and Voekst (2001) found that the impact of consumer confidence on consumer spending was stronger compared to the US market. Cotsosmitis and Kwan (2006) investigated the subject matter for Belgium, Denmark, France, Italy, Germany, Portugal, Spain, the Netherlands, and the United Kingdom. They examined whether the aggregate Consumer Confidence Index (CCI) for the EU had an impact on consumer spending in each selected country and whether each individual country's confidence index had an impact on the consumer spending of the corresponding country. Their results showed that six out of nine countries' consumer spending was affected by aggregate CCI, but the impact of the individual country's CCI on the corresponding country's consumer spending was generally small and could be explained away by other macroeconomic controlling variables.

Some recent research has examined whether a positive relationship can be found when both customer satisfaction and spending are aggregated across the US population. Fornell et al. (2010) viewed the role of customer satisfaction as a predictor of discretionary spending growth and found that satisfaction from previous purchases could explain 23% of the variation in the next quarter's spending growth. Ramasamy and Yeung (2010) further confirmed and extended the findings of Fornell et al. (2010). They examined the influence of the ACSI and CCI on Personal Consumption Expenditure (PCE) and concluded that compared to the CCI, the ACSI can be used to predict a wider range of consumption categories. The findings from Fornell et al. (2010) and Ramasamy and Yeung (2010) provided empirical evidence to support the plausible relationship between customer satisfaction and aggregate consumer spending. As both studies were conducted based on the US economy, the authors have called for further investigation to affirm the findings in other regions. The CCI reported in some European countries therefore provides a good opportunity for this investigation. Considering the mature market economy and sophisticated institutional structure in Europe, it is reasonable for us to develop the first hypothesis:

**H1. In the selected European countries, aggregate customer satisfaction is a significant driver of consumer expenditure.**

### 4. Cross-country moderating effects

It is noted that customer satisfaction may not yield an invariable influence on consumer expenditure and that its impact will be contingent on certain conditions. For instance, Fornell et al. (2010) suggested that households’ Debt Service Ratio (DSR) might moderate the impact of consumer satisfaction. As rising debt levels may restrain consumers' abilities to pay, the ability of customer satisfaction to predict future consumption will be attenuated. Ramasamy and Yeung (2010) examined this issue by taking the perspective of product difference. They found that the forecasting capacity of the ACSI had a significant impact on the consumption of durable goods and services, compared with the CCI, which had a significant effect on non-durables.

In this paper, we extend our understanding of the relationship between customer satisfaction and consumption expenditure by investigating country-specific issues and their impact on this link. In a study investigating the cross-national determinants of customer satisfaction, Morgeson et al. (2011) alluded to a range of country-level differences that “influence how consumers perceive and respond to their consumption experience and the level of satisfaction delivered by an economy” (p. 200). These differences include cultural, politico-economic, and socio-economic factors. In an earlier study, Ogikubo et al. (2009) compared the influence of economic conditions and economic growth on customer satisfaction in the US, Japan and Sweden and concluded that economic, socio-political and cultural differences account for the variation in customer satisfaction among these countries. When studying the relationship between customer satisfaction and repurchase at the individual level, Seiders et al. (2005) also argued that the factors that moderate the relationship could be summarized into three categories: consumer, relational, and marketplace characteristics. Some of these characteristics, such as income, education, and market competition, are also relevant at the aggregate level and can be incorporated into the socio- and politico-economic domain. Therefore, building on the factors considered in these previous studies, we propose four main factors that may moderate the influence of customer satisfaction on consumption expenditure: economic structure, culture, political economy and social economy. A visual representation of our framework is shown in **Fig. 1**, and the ensuing hypotheses are developed to represent each of these factors.

A notable difference among the economic structures of countries is the proportion of the national GDP represented by the service sector. Services are generally characterized as being highly dependent on personnel and customization to suit heterogeneous needs in creating customer satisfaction (Anderson, Fornell, & Rust, 1997). Unlike physical goods, services are intangible, making them hard to measure, store, and test (Grönroos, 1990). If something goes wrong in the service process, it is often difficult to identify the reasons and implement quality control (Hoffman & Bateson, 1997). These characteristics help explain why it is challenging to achieve a high level of service performance (Fornell & Johnson, 1993). Compared to purchasing goods, where consumers can often assess product quality prior to purchase through free samples, test-drives and manuals, it is more difficult for consumers to evaluate and understand service offers (Edwardsson, Johnson, Gustafsson, & Strandvik, 2000). Even if the service concept is clear and the prerequisites for a quality service are in place, the service’s intangibility makes it difficult for the service firm to display or communicate the offering in the marketplace (Grönroos, 1990). This situation results in customers placing great emphasis on their prior experience and experience-based referrals from others. Therefore, it could be argued that customer satisfaction should carry a greater weightage for consumer spending on services than on goods. Thus, we propose the following:

**H2. Customer satisfaction should be a more significant driver of consumer expenditure in countries where the service sector is a larger component of the economy.**

National culture is another country-specific feature that is worth considering. Hofstede defined culture as “the collective programming of the mind which distinguishes the members of one group or category of people from those of another” (1994, p. 4). Such “collective programming” often influences the member’s values, beliefs and norms (Pizam, Pine, Mok, & Shin, 1997) and their role as consumers. Prior research has confirmed the importance of culture to a range of customer satisfaction issues, such as the link between culture and consumer expectations (Donthu & Yoo, 1998), complaint behavior (Liu & McClure, 2001) and the willingness to report dissatisfaction (Crots & Erdmann, 2000).

Despite the popularity of Hofstede’s five cultural dimensions in cross-cultural studies, it is not viable for us to adopt those dimensions in this study because two of our sample countries lack sufficient data coverage in those areas. Following Inglehart and Baker (2000), we use two dichotomous measures as national cultural indicators: survival vs. self-expression values and traditional vs. secular-rational values. We map the countries selected in this study on those dimensions, using data from the 2000 wave of the World Values Survey (Inglehart & Welzel, 2005). As shown in **Fig. 2**, the survival vs. self-expression
dimension (horizontal) is able to explain some differences between our sample countries far better than the traditional vs. secular-rational dimension (vertical). This observation leads us to focus on the former in the analysis. The survival/self-expression dimension is linked to the transition from an industrial society to a post-industrial society. Inglehart and Baker (2000) explained that societies with strong survival values tend to emphasize economic and physical security, show relatively low levels of subjective wellbeing, report relatively poor health, and demonstrate low interpersonal trust. Societies with self-expression values tend to show the opposite results on these issues.

It is argued that the survival/self-expression value likely influences how consumers perceive and respond to their consumption experience and the level of satisfaction delivered within an economy. For instance, Morgeson et al. (2011) found that consumers in self-expressive societies will express higher levels of satisfaction than those in societies with survival values. The higher levels of interpersonal trust developed in self-expressive societies result in more active consumer interaction, making positive word-of-mouth an important driver of purchase. In addition, the emphasis on subjective wellbeing and quality of life makes consumers more willing to pay for a satisfied experience. In contrast, the influence of experience may be less significant in survival societies due to the lower level of consumer interaction and the emphasis on economic and physical security. For these reasons, we suggest the following:

**H3.** Customer satisfaction should be a more significant driver of consumer expenditure in self-expressive societies than in survival societies.

In our study, we also attempt to examine the moderating effect of two socio-economic factors: education and income. These moderating factors are important because they can explain cross-country variation in the satisfaction–consumption relationship at the national level. Furthermore, these variables can be influenced by policy-makers.

Previous study suggests that the variation among the satisfaction rating can be explained by several consumer characteristics. For instance, Bryant and Cha (1996) analyzed the ACSI and found that consistent differences in satisfaction levels do exist among different socio-economic groups, irrespective of the types of goods. The study also demonstrates a negative relationship between customer satisfaction and the consumer’s level of education. In other words, customer satisfaction declines as education level rises. However, they could not ascertain the extent to which such differences in ratings can translate into repurchase behavior. Whether higher satisfaction reported by a relatively less-educated consumer will lead to a higher chance of repurchase is still an unanswered question.

Mittal and Kamakura (2001) are among the first to examine the moderating effect of consumer characteristics (e.g., age, gender, and education) on the satisfaction–repurchase link. They found that for the same level of rated satisfaction, consumers with more education tend to show a lower probability to repurchase than those with less education. The authors argued that “consumers with higher education could have greater ability to search and are cognizant of superior alternatives in the market” (p. 139). In another study, Capraro, Broniarczyk, and Srivastava (2003) investigated the influence of consumer knowledge on the relationship between customer satisfaction and customer defection. Their findings suggested that the likelihood of customer defection will increase as consumers obtain more knowledge and information about the alternatives, making customer satisfaction a weaker predictor of repurchase. Given these established links at the individual level, it is reasonable to argue that such relationships may also apply at the aggregate level. We argue that as levels of education increase, consumers’ ability to effectively search and evaluate information surrounding products/services will also increase. As consumers increase their reliance on external information about products/services, their satisfactory experiences tend to play a relatively less important role in determining their purchases. Thus, we propose the following hypothesis:

**H4.** Customer satisfaction becomes a less significant driver for consumer expenditure as the nation’s level of education increases.

Previous research also examines the influence of income on consumer satisfaction. Some studies suggest that as income increases, consumers tend to become more critical of the goods they consume and harder to please (Anderson, Pears, & Widener, 2008; Bryant & Cha, 1996). Morgeson et al. (2011) further investigated this influence at the national level and found that consumers in societies with a higher
per capita income tend to express lower satisfaction with goods and services. The authors explain that “consumers, as their wealth grows with the nation’s economy (over long periods of time), gradually become more demanding” (p. 212).

However, in regard to the satisfaction–repurchase link, it is postulated that income should intensify the relationship between satisfaction and repurchase behavior. For instance, by measuring objective repurchase behavior, Seiders et al. (2005) show that household income will positively enhance the effect of satisfaction on repurchase visits and spending. Fornell et al. (2010) also argue that if consumers’ discretionary spending shrinks, it will attenuate the impact of a consumer’s satisfaction with prior purchases. This reasoning implies that customer satisfaction can be a better predictor for consumers with higher income than their poorer counterparts. Lower-income consumers are generally more price-driven in their purchases and more subject to the influence of price discounts and in-store promotions. In such situations, it is reasonable to assume that lower-income consumers place a higher value on price and that past experiences become less important in influencing the purchase decision. Higher-income consumers tend to pay more attention to the hedonic and functional utility of their purchases and less attention to price. Thus, their past satisfactory experiences become more important to their subsequent purchases. If this logic exists at the individual level, it is reasonable to make a similar argument at the aggregate level. Thus, we posit that:

H5. Customer satisfaction becomes a more significant driver of consumer expenditure as the nation’s per capita income increases.

Miller and Kim (2012) define an economically free society as one where “the power of economic decision-making is widely dispersed, and the allocation of resources for production and consumption is on the basis of free and open competition so that every individual or firm has a fair chance to succeed.” The potential link between market characteristics and customer satisfaction has long been noted (Johnson, Herrmann, & Gustafsson, 2002; Seiders et al., 2005). It is expected that consumers should experience greater utility/satisfaction in a free market. For instance, Johnson et al. (2002) argue that firms in countries with higher levels of economic freedom will have greater motivation to satisfy their customers. In a more recent study, Morgeson et al. (2011) also found a positive relationship between economic freedom and customer satisfaction. They argued that increased openness to international commerce and internal business will “broaden the number, quality and pricing of competitive alternatives in a manner beneficial to the consumers (and their satisfaction)” (p. 202).

Aside from its direct impact on customer satisfaction, economic freedom also plays a positive moderating role on customer satisfaction and repurchase behavior (Seiders et al., 2005). An economy with lower economic freedom tends to have fewer firms and thus fewer choices available to consumers. Therefore, a consumer may continue to purchase despite low satisfaction because of the limited choices in the marketplace. This situation will make customer satisfaction a weaker predictor of consumer spending. In countries with greater economic freedom, firms will show a stronger willingness to establish customer satisfaction for fear that unsatisfied consumers will switch to other choices. Therefore, firms will be motivated to provide various economic incentives and relationship programs to customers, which will help “enhance the positive effect of satisfaction on repurchase behavior” (Seiders et al., 2005, p. 31). Thus, we posit that:

H6. Customer satisfaction is a relatively more significant driver of consumer expenditure in countries with higher economic freedom.

5. Data, methodology, and results

Our study employs panel data modeling techniques. Panel data models do everything that is possible with a time series model while also controlling for individual-specific, time-invariant variables and addressing unobserved heterogeneity. A few advantages of panel data are particularly important to the present study. First, the estimation can take country-specific heterogeneity into account explicitly; second, panel data, by nature, provide more variability and less collinearity among variables and thus increase estimation efficiency; and third, panel data can minimize the bias that might result if we aggregate countries into broad groupings. The use of panel data in estimating common relationships across countries is particularly appropriate for the current research agenda because it allows the identification of country-specific effects that control for missing or unobserved variables (Judson & Owen, 1997). In other words, it assumes that time-invariant country differences exist, the effects of which could impact or bias the outcome variable (consumption in our case). Thus, the use of panel data models allows us to control for these time-invariant variables. Most importantly, the country-specific estimation of the relationship between customer satisfaction and consumption at the economy level demands a time series of significant length. Until time series data of significant length are available, the only feasible alternative for conducting the estimation is to employ panel data models and techniques, i.e., pooling a sample of several countries over a certain period of time to conduct such estimations so as to obtain a common estimation.

Spurious regression is less likely when modeling with panel data compared with time series data, especially for models with level data. Phillips and Moon (1999) and Kao (1999) explained that this independence works effectively by smoothing out the usual unit root dependency for each unit. Thus, the technique helps mitigate the spurious regression problem arising in the time series case.

Data for the analysis were collected from five data sources: Euromonitor; the European Commission’s consumer surveys; Jan Eklöf’s EPSI (European Performance Satisfaction Index); the World Values Survey; and the World Bank’s World Development Indicators. Our data analysis covers the period from 1999 to 2011 in nine countries: the Czech Republic, Denmark, Estonia, Finland, Greece, Latvia, Lithuania, Portugal, and Sweden. The selection of countries was based on the availability of all required data from a common source for at least a three-year period.

The EPSI, formerly known as the European Customer Satisfaction Index (ECSI), is based on a microeconomic model that considers causal relationships among a set of antecedents of customer satisfaction. It considers customer satisfaction to be a cumulative experience rather than the result of recent transactions and has been built to be compatible with other national satisfaction barometers (Eklöf et al., 1999). The EPSI is currently under the management of the European Foundation for Quality Management, the European organization for quality, and the academic network, International Foundation for Customer Focus (see Eklöf & Selivanova, 2008; Eskildsen & Kristensen, 2007 for more description and methodology of EPSI). Eklöf and Westlund (2002) suggested that the EPSI is based on a thorough analysis of theory and an implementation of best practice methodology of data collection, measurement and analysis. It uses survey data, collected by telephone interviews, to create latent variables, e.g., customer expectations, perceived product quality, perceived service quality, perceived value and corporate image, to compute the customer satisfaction measurement. Moreover, it is a well-structured method to measure customer satisfaction and represents another variation to the ACSI model. Indeed, both the EPSI and the ACSI were modified from the first uniform national measurement instrument for customer satisfaction and customer loyalty in Sweden—the Swedish Customer Satisfaction Barometer.

The first annual wave of the EPSI survey (1999) conducted more than 100,000 interviews across 11 countries; some 10 industries were surveyed, including retail banking, fixed and mobile line telecommunications, and supermarkets. These industries were common across countries, allowing for regional aggregation. In 2011, the survey expanded to 1,000,000 respondents across 16 countries. In most countries, the EPSI covers the financial sector (banking and insurance), the ICT sector
The dependent variable (PCE). According to Persyn and Westerlund (2008), the two tests are designed to pool information over all the cross-sectional units to test the null of no co-integration for all cross-sectional units against the alternative of co-integration for all cross-sectional units. Rejection of the null should therefore be taken as evidence of co-integration for the panel as a whole. Thus, they can be regarded as stringent tests. P1 sets the number of leads and lags in the test equation based on the number of years in the data, while P2 sets the number of leads and lags in the test based on the Akaike information criterion. Our results show that PCE and PDI are co-integrated, but no long-run co-integration relationship was found between PCE and CSI. We use these results to guide the rest of the model building.

Given that PCE and PDI are fundamentally and statistically co-integrated, we started the modeling by fitting the long-run consumption model given in Eq. (1) with panel data and a fixed effect specification, as suggested by the relevant Hausman test (see Abeysinghe & Choy, 2004; Davidson, Hendry, Sibja, & Yeo, 1978; Saad, 2011):

\[
PCE_{it} = (\alpha_i + u_i) + \beta_1 PDI_{it} + \epsilon_{it}
\]

where \( \alpha_i \) (i = 1 \ldots 9) is the unknown intercept for each country; \( u_i \) is the error term; subscript i denotes country and t refers to time. Next, given the unit root and co-integration results, the short-run dynamics of consumer satisfaction (\( \Delta CSI \)), consumer confidence (\( CSI \)) and other controlling variables are appended to the consumption model. We follow Fornell et al. (2010) and use inflation (\( \Delta CPI \)) and debt (\( DEBT \)) as controlling variables. The expected signs for the controlling variables are negative and positive, respectively. Given the additional variables, the following difference equation is fitted to guide the rest of the analysis:

\[
\Delta PCE_{it} = (\nu_1 + \epsilon_{it}) + \beta_2 \Delta PDI_{it-1} + \beta_3 ECT_{it-1} + \delta \Delta CSI_{it-1} + \theta CPI_{it-1} + \beta_4 \Delta CPI_{it-1} + \beta_5 DEBT_{it-1}
\]

where \( \nu_1 \) (i = 1 \ldots 9) is the unknown intercept for each country; \( \epsilon_{it} \) is the error term; \( \Delta PCE \) and \( \Delta PDI \) are the changes in personal consumption expenditure (per capita) and personal disposable income (per capita), respectively. ECT is the error-correction term obtained from Eq. (1) (i.e., \( PCE_{it-1} - \beta_0 PDI_{it-1} \)). \( \Delta CSI \) is the change in customer satisfaction, \( CSI \) is the level of consumer confidence, and \( \Delta CPI \) and \( DEBT \) are as previously defined.

All variables in Eq. (2) are stationary. Moreover, it is believed that Eq. (2) expresses the consumption function correctly. Spanos (1989) provides a comprehensive analysis concerning different forms of the consumption function. The theoretical and statistical adequacies of alternative models were compared to judge the extent to which the selected models conform to theory and fit the theoretical parameters

---

4. Consumer confidence was found to be I(0), and not consistent with Lemmens, Croux, and Dekimpe (2007). This may be due to differences in data and methodology used, sample period and countries selected. Our analysis was based on (computed) yearly panel data covering the period from 1999 to 2011 and panel data techniques that allowed to increase the power of unit root tests based on a single time series, whereas Lemmens et al’s analysis was based on monthly data from Nov 1995 to Feb 2004. Furthermore, their unit root tests were carried out separately for each selected country. Additionally, Lemmens et al., focused on large EU economies, whereas our sample comprises relatively smaller EU economies.

5. The lagged consumption growth was considered as a RHS variable but the inclusion of it turns the error-correction term positive, that is, inconsistent with consumption theories.
derived by consumption theories. Our Eq. (2) is identical to Model 8 in Spanos (1989). The same model proved empirically successful for both the US and the UK.

Some diagnostic tests are used to guide our selection of the estimator. We use Breusch and Pagan’s (1980) LM test to test the null hypothesis of no random individual effects (both the US and the UK. Spanos (1989). The same model proved empirically successful for that all regressors are statistically significant. Therefore, the FE estimator should be the best alternative among the panel estimators, we use the Hausman test to decide on the choice of appropriate, and the panel estimator should be adopted (see Baltagi & Griffin, 1983; Costantini & Martini, 2010; Shaanan, 1997). Among the panel estimators, we use the Hausman test to decide on the choice of fixed effect (FE) versus random effect (RE) estimation. Under the null hypothesis, both estimations are consistent. In contrast, the alternative hypothesis suggests FE over RE. The test statistic is 15.9 (p = 0.014). Therefore, the FE estimator should be the best alternative among the three competing estimators (OLS, FE and RE).

The final estimations are shown in Table 2. The fitted model suggests that all regressors are statistically significant at the 5% level. Based on the principle that the effects of independent variables are solely within-cluster effects in a fixed effects model (Bartels, 2002; Wooldridge, 2002), for a given country, as ΔCSI increases across time by one unit, PCE per capita increases by 57.84 units, holding fixed the controlling variables and country-specific effects. The size of the CCI coefficient is more than three times smaller than ΔCSI. The variables entered explain 56% of variations in ΔPCE. The Jarque–Bera test statistic does not reject the null hypothesis that errors are normally distributed. Note that the reported intercepts are the average of country-specific intercepts. However, in a panel data model, the country-specific intercepts have weak explanatory value (Wooldridge, 2002).

To confirm the predictability and examine the lag structure of ΔCSI, we follow Fan and Wong (1998), and Ramasamy and Yeung’s (2010) methodology to check the bivariate relationship between ΔPCE and ΔCSI. We regress ΔPCE on different lags of ΔCSI and consider the changes in the relationship when accounting for the effect of the main controlling variable. Previous researchers concluded that the independent variable is a good predictor when different lag structures do not distort the significance of specific independent variables. The results are reported in Table 3.

On the LHS of Table 3, ΔPCE is regressed on ΔCSI and lagged ΔCSIs. One to three lags are considered. The coefficient(s) of ΔCSI are (jointly) significant for all three cases. Similarly, on the RHS of Table 3, the same result is revealed when the controlling variable and its lags are added. Given these results, we can regard CSI as a good predictor of PCE, and its effects remain significant even when more lags were added. Therefore, H1 is not rejected.

Table 2
Estimation results of total sample.

<table>
<thead>
<tr>
<th>ΔPCEit</th>
<th>[\Delta PCE_{it} = \alpha + \beta_1 \Delta CPI_{it-1} + \beta_2 \Delta DEBT_{it-1} + \beta_3 \Delta CSI_{it-1} + \beta_4 MFI_{it-1} + \beta_5 \Delta CSI_{it-2} + \beta_6 \Delta CSI_{it-3} + \epsilon_{it}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>[R^2 = 0.58; Durbin Watson statistic = 1.89; Jarque-Bera statistic = 0.24; N = 71]</td>
</tr>
</tbody>
</table>

To examine hypotheses H2–H6, the proposed moderating factors are added to Eq. (2) one at a time. Consider Eq. (3):

\[\Delta PCE_{it} = (\alpha_i + u_{it}) + \beta_1 \Delta CPI_{it-1} + \beta_2 \Delta DEBT_{it-1} + \delta \Delta CSI_{it-1} + \theta \Delta CPI_{it-1} + \beta_3 MFI_{it-1} + \beta_4 MFI_{it-1} \times \Delta CSI_{it-1},\]

where MFI_{it-1} is the moderating factor and MFI_{it-1} × ΔCSI_{it-1} is the interacting variable; the rest of the variables were previously defined. Given the hypotheses we developed, five moderating factors are considered one at a time. These factors and their sources of data are listed and described in Table 4.

Except for ΔPDI, the selected moderating factors are either time-invariant or sluggish (rarely changing) variables. Neither the fixed- nor random effect estimators are feasible options for handling such estimations. The former would sweep away the moderating variables when eliminating the country-specific effects; and the latter would impose the assumption that the entity’s error term was not correlated with any RHS variables. We showed that this assumption was restrictive for our data when estimating Eq. (2). To fit Eq. (3) with a time-invariant regressor, Plumper and Troeger’s (2007) Panel Fixed Effects Regression with Vector Decomposition approach is used. The proposed estimator allows us to estimate time-invariant or rarely changing variables with a fixed effect model. Using Monte-Carlo experiments, Plumper and Troeger (2007) showed that their proposed processes performed better than other methods for handling panel data estimations with time-invariant regressors. The estimations of Eq. (3) with one moderating factor entered at a time are reported in Table 5.

An interaction term is a way to show that the predictive effect varies by subgroups of predictors. Focusing on the interaction terms, our results in Table 5 suggest that SURV negatively moderates, while SERV, EDU, and FREE positively moderate the relationship between changes in customer satisfaction and changes in consumption. In the empirical analysis, we provided supportive evidence for H2 and H6 at the 10% significance level and for H4 at the 5% significance level. However, we did not find supportive evidence for H3 and H5. In H3, we argued, based on previous literature, that customer satisfaction should positively affect consumption expenditure in societies that are self-expressive. However, our results find the interacting variable to have a negative coefficient (significant at the 5% level). This observation suggests that the satisfaction–consumption link is actually weaker in countries that are high in self-expressive values.

### Table 3
Regressing ΔPCE on lagged ΔCSI and lagged ΔPDI.

<table>
<thead>
<tr>
<th>RMSE of the baseline model</th>
<th>ΔPCE = [\sum_{k=1}^{3} \delta_k \Delta PDI_{it-k} + \sum_{k=1}^{2} \gamma_k \Delta CSI_{it-k} + \epsilon_{it}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k = 1)</td>
<td>RMSE = 0.615.33 (\delta_1 = 0.71)</td>
</tr>
<tr>
<td>(k = 2)</td>
<td>RMSE = 0.645.07 (\delta_1 = 0.71)</td>
</tr>
<tr>
<td>(k = 3)</td>
<td>RMSE = 0.674.22 (\delta_1 = 0.71)</td>
</tr>
</tbody>
</table>

6. Discussion and implications

Our study is an attempt to establish a satisfaction–outcome link at the macro level. Using well-established econometric techniques and models, we find that customer satisfaction plays a crucial role in determining consumption expenditure in selected European countries.

---

8 We follow procedures for the detection of interaction terms developed by previous researchers like Cohen and Cohen (1983) and Cox (1984). These procedures examine the significance of interaction terms one at a time to avoid over-fitting. Including all interacting terms into one equation would also result in a co-linearity issue, as ΔCSI_{it-1} is a product term in all the interaction terms. A test of co-linearity of the moderating factors also produced negative results.
From a theoretical point of view, the paper adds to the scarce but growing literature on customer satisfaction and consumer expenditure at the macro level. We have extended the findings of Fornell et al. (2010) and Ramasamy and Yeung (2010) to another important economic region. In addition, we find that cross-country differences, including economic structure, culture, and socio- and political-economic factors moderate the satisfaction—consumption relationship. Specifically, we find that the influence of customer satisfaction is stronger in economies with a high reliance on the service sector, as we hypothesized in H2. While prior research has investigated the link between customer satisfaction and business performance across different industries, these studies tend to be conducted within a country, mostly in Sweden (Anderson et al., 1997; Edvardsson et al., 2000; Nilsson, Johanson, & Gustafsson, 2001). Our findings, to some degree, are in line with Edvardsson et al. (2000), who argued that customer satisfaction shows a more positive impact on revenue growth for services than for products. However, our findings provide even more convincing evidence, with data covering nine countries over a 12-year period.

Our findings suggest that a nation’s level of education and its economic freedom are two important moderators of the customer satisfaction—consumption link. In H4, we suggested that the link between customer satisfaction and consumption expenditure is weaker among countries with higher levels of education. Because EDU denotes the percentage of the labor force with only primary education, a higher number indicates a lower level of education. Our results of a significant coefficient (60.254) suggest that customer satisfaction affects consumption expenditures more in countries where the level of education is relatively lower. Previous research has confirmed that higher levels of education are associated with lower levels of loyalty (Mittal & Kamakura, 2001). Highly educated customers tend to demand more information related to their purchase, and therefore, they affect the degree of sensitivity with which investments in satisfaction convert to loyalty (Bae, Russell, & Rego, 2011). This finding implies that consumers’ increasing capacity to search for and process information, which results from an increase in education, will weaken their reliance on previous consumption experiences in future purchases. Thus, in countries where levels of education are higher, converting satisfaction into increased consumption is bound to be more challenging.

Our results also show that the satisfaction—consumption link is stronger for countries with higher economic freedom, as argued in H6. This outcome is consistent with the study of Jones and Sasser (1995), who showed that the satisfaction—loyalty link is more prevalent in markets where competition is intense. As explained earlier, firms in economies with greater freedom tend to place high value on customer satisfaction and are motivated to retain customers and encourage repeat purchases through various incentives and relationship programs. Our results show that the satisfaction—loyalty link does extend to increased consumption.

Nevertheless, our findings show some unexpected results regarding the influence of culture on the satisfaction—consumption link. In H3, we argued that customer satisfaction should be a more significant driver for consumption expenditure in self-expressive societies than survival ones. However, our results suggest that the satisfaction—consumption link is weaker for countries that are high on self-expressive values. An in-depth investigation into this issue shows that societies with strong survival values tend to emphasize collectivism and those with strong self-expression values emphasize individualism (Inglehart & Oyserman, 2004). Liu, Furrer, and Sudharshan (2001) found that collectivist consumers tend to stick to the same service provider once they are satisfied. This result is further endorsed by Jin, Park, and Kim (2008), who also found a stronger satisfaction—loyalty link in Korea (a highly collectivized society) than the US (a high individualistic society). Compared with self-expressive societies, customers in societies with high survival values pay more attention to fundamental economic and physical security (Inglehart & Baker, 2004) where product/service quality is less guaranteed. Therefore, consumers tend to rely on past satisfactory experiences as a means to avoid uncertain quality in future consumption. This situation could explain why the link between satisfaction and consumption is relatively more significant in societies with stronger survival values.

### Table 4

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Modifying factor</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERV</td>
<td>Proportion of the services sector in GDP</td>
<td>Proxy for the differences in terms of economic structure.</td>
<td>World Bank’s World Development Indicators</td>
</tr>
<tr>
<td>PDI</td>
<td>Income per capita</td>
<td>Proxy for socio-economic differences. Disposable income measured as gross income minus social security contributions and income tax, in constant US dollars (in per capita values).</td>
<td>World Bank’s World Development Indicators</td>
</tr>
<tr>
<td>EDU</td>
<td>Education levels</td>
<td>Proxy for socio-economic differences. Measured by the proportion of the country’s labor force with only primary education.</td>
<td>World Bank’s World Development Indicators</td>
</tr>
<tr>
<td>SURV</td>
<td>Survival values</td>
<td>Proxy for cultural differences among countries. Measured based on a battery of questions as per Inglehart &amp; Welzel, 2005.</td>
<td>World Values Survey</td>
</tr>
<tr>
<td>FREE</td>
<td>Economic freedom</td>
<td>Proxy for political-economic differences. Similar to Morgeson et al. (2011), we use trade and business freedom to represent economic freedom.</td>
<td>Heritage Foundation</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>MF = variable</th>
<th>APDL</th>
<th>SERV</th>
<th>FREE</th>
<th>SURV</th>
<th>EDU</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1628.991</td>
<td>622.140</td>
<td>101.955</td>
<td>129.044</td>
<td>98.541</td>
</tr>
<tr>
<td>$\Delta PDL_{it} - 1$</td>
<td>0.378</td>
<td>0.000</td>
<td>0.340</td>
<td>0.373</td>
<td>0.005</td>
</tr>
<tr>
<td>$ECS_{i,k} - 1$</td>
<td>0.034</td>
<td>0.000</td>
<td>0.015</td>
<td>0.009</td>
<td>0.003</td>
</tr>
<tr>
<td>$\Delta CS_{it} - 1$</td>
<td>0.430</td>
<td>0.011</td>
<td>0.011</td>
<td>0.008</td>
<td>0.003</td>
</tr>
<tr>
<td>$CC_{i,k} - 1$</td>
<td>0.043</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Delta CC_{it} - 1$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$DEBT_{it} - 1$</td>
<td>0.317</td>
<td>0.079</td>
<td>0.077</td>
<td>0.077</td>
<td>0.003</td>
</tr>
<tr>
<td>$\Delta DEBT_{it} - 1$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$FR_{i,k}$</td>
<td>0.098</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$fm_{i,k}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.538</td>
<td>0.480</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
</tr>
</tbody>
</table>

* ** and * denote significance at 1%, 5% and 10%.
Our study offers some helpful managerial and policy implications. At the managerial level, our findings elevate the role of marketers and marketing activities in customer satisfaction to one of national interest. The marketing effort emphasizes better ways to deliver goods and services to create more satisfying experiences. Though we do not directly measure how marketing activities contribute to the economy, the empirical evidence in the study clearly confirms the economic contribution of customer satisfaction, which is only achieved through good marketing. Thus, marketers need to understand that their efforts to satisfy the appetites of customers do not merely meet sales and profit targets at the firm level but also have a direct impact on economic growth. The marketing effort to improve customer satisfaction will help increase the aggregate level of consumer spending, which ultimately benefits both the company and economy.

The findings imply that companies should not trade customer satisfaction for a short-term sales target because declining customer satisfaction will hurt both the company and industry growth in the long run. The findings might also imply that dramatically reducing marketing budgets in reaction to economic downturns is a questionable strategy for companies. As marketing managers lose the human and other resources necessary to take care of customers, the resultant decline in customer satisfaction may lead to a vicious circle. Managers would further benefit from a better understanding of moderating variables, such as education and culture, which can be used to segment customers. As Multinational Corporations (MNCs) are operating in many countries, examining cross-national differences in this customer satisfaction–consumption link has potential value. Our results suggest that MNCs should realize that companies’ efforts to satisfy customers will have different results in different countries. These macro-level factors also impact the reward of their efforts beyond their product and service performance. The firm can choose the proper markets where their efforts to develop customer satisfaction will yield more consumption.

The implications for policy-makers are even more direct. Our findings show the need for every country to establish some form of CSI. Policy-makers should understand that such a national index not only gauges the quality of goods and services consumed but also provides a valid forecast for future consumption activities at the macro level. Therefore, as a matter of national interest, policy-makers should provide incentives and increase the pressure on firms to constantly invest in customer satisfaction. This is especially true for economies with a strong service sector because consumer spending is influenced more strongly by customer satisfaction in the service sector. Services often score very low in customer satisfaction due to the characteristics explained earlier. Even in the US, where customer satisfaction is higher than in European countries, e.g., Germany and Sweden (Johnson et al., 2002), the bottom five industries in the ACSI are consistently industries in the service sector (e.g., airlines, television and telephone services). As services become even more dominant, policy-makers should give increasing attention to customer satisfaction in this sector. As advanced economies struggle with low economic growth rates, policy-makers may leverage the lagged customer satisfaction in services to encourage economic growth. In addition, our findings on economic freedom suggest to policy-makers that greater business freedom will further enhance the role of customer satisfaction in boosting long-term economic growth.

The findings of this study are limited by the availability of both time series and cross-sectional data. In terms of time series data, the EPSI is only collected on an annual basis, which limits the variation along the time dimension. As for cross-sectional data, several significant countries are not within the EPSI network including Germany, France and the UK. The small number of countries selected in the study also limits our empirical strategy to evaluate other cross-country differences and generalize our findings across the entire European continent. As for the EPSI measure itself, Steenkamp and Baumgartner (1998), Baumgartner and Steenkamp (2001) and Harzing et al. (2009) warned that a similar measure might not necessarily be exactly equivalent across countries, largely due to the differences in the response styles among cultural groups. It is also likely that some cultural groups are harder to please than others. Remedial methods to control for various response biases have been proposed, but these methods are more relevant for conducting research with primary data. However, such potential biases may be assumed to be minimal for the current study because our model used changes in satisfaction. Furthermore, the EPSI has been widely regarded as an exercise based on a generalizable conceptual framework. Given these data limitations, future research could validate our results by including more European countries as CSI data become available. In addition, further research into those marketing activities that relate to customer satisfaction could be undertaken. For instance, the impact of advertising on customer satisfaction at the macro level could be evaluated to understand the real role of the former in improving macroeconomic performance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jresmar.2013.06.001.

References


See BBC News (2003) for Pearson cutting the marketing budget during the economic downturns; see Qualifed Remodeler Survey (2005) for construction firms setting marketing budget as the first-to-cut item; see Sweeney (2001) for high-tech industry slashing marketing budgets during economic downturns.


