Understanding and managing international growth of new products

Stefan Stremersch\textsuperscript{a,b,*}, Gerard J. Tellis\textsuperscript{c}

\textsuperscript{a}Department of Marketing, School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands
\textsuperscript{b}Goizueta Business School, Emory University, Atlanta, GA 30322-2710, USA
\textsuperscript{c}Marshall School of Business, University of Southern California, Los Angeles, CA 90089-1421, USA

Received 12 August 2003; received in revised form 19 January 2004; accepted 9 July 2004

Abstract

Growth is one of the most compelling goals of managers today. This paper addresses the following questions about the international growth of new products in Europe: Does the pattern of growth differ across countries? If so, does culture or economics explain the differences? What are the implications of these results for new product strategy?

The results show that the pattern of growth differs substantially across European countries. These differences are explained mostly by economic wealth and not by culture. The study addresses the implications of these results for: (a) the choice of a waterfall versus sprinkler strategy for the introduction of a new product; (b) the global versus local marketing of a new product; and (c) managing a firm’s expectations about new product growth.

© 2004 Elsevier B.V. All rights reserved.

Keywords: New product growth; International marketing; International diffusion

1. Introduction

Growth is one of the most persistent and compelling goals of managers today. Firms’ accountability to stock markets or profit-seeking owners ensures that next to profitability, growth is the most important goal of ongoing enterprises. Some analysts rank growth as an even higher goal than profitability because of its potential for future revenues and profits.

For most firms, the introduction of new products is the primary engine of growth. However, new consumer durables do not grow evenly from the instant of introduction. Rather, they typically show an S-shaped sales curve consisting of at least three distinct stages: (1) an introductory stage of little or no growth; (2) a growth stage with very high growth; and (3) a maturity stage marked by little or negative growth (Mahajan, Muller, & Bass, 1990). The growth stage is bounded by what has been
Table 1: Overview of prior related literature

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Independent variables</th>
<th>Sample composition</th>
<th>Countries</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gatignon et al. (1989)</td>
<td>$p$ and $q$ (Bass model)</td>
<td>Cosmopolitanism, mobility, sex roles</td>
<td>Dishwasher, deep freezer, lawn mower, pocket calculator, car radio, color TV</td>
<td>Belgium, Denmark, France, W. Germany, Italy, Netherlands, UK, Austria, Finland, Norway, Portugal, Spain, Sweden, Switzerland</td>
</tr>
<tr>
<td>Takada and Jain (1991)</td>
<td>$q$ (Bass model)</td>
<td>Culture (high vs. low context), communication (homophilous vs. heterophilous)</td>
<td>Black and white TV, washing machine, air conditioner, car, refrigerator, calculator, vacuum cleaner, radio</td>
<td>US, Japan, South Korea, Taiwan</td>
</tr>
<tr>
<td>Helsen et al. (1993)</td>
<td>$p$, $q$, and $s$* (Bass model)</td>
<td>Mobility, health, foreign trade, standard of living, cosmopolitanism</td>
<td>Color TV, VCR, CD player</td>
<td>Austria, Belgium, Denmark, Finland, France, Japan, Netherlands, Norway, Sweden, Switzerland, UK, US</td>
</tr>
<tr>
<td>Kalish et al. (1995)</td>
<td>Cumulative adopters, profitability</td>
<td>Lead/lag effect, costs, competition, length of life cycle, market size, innovativeness, market growth</td>
<td>Analytical model</td>
<td>Analytical model</td>
</tr>
<tr>
<td>Putsis et al. (1997)</td>
<td>Cumulative adopters, sales</td>
<td>Cumulative adopters in other countries (cross-country effects), population, TV use, GDP per capita</td>
<td>VCR, CD player, microwave oven, home computer</td>
<td>Great Britain, Germany, France, Italy, Spain, Belgium, Denmark, Netherlands, Austria, Sweden</td>
</tr>
<tr>
<td>Dekimpe et al. (2000)</td>
<td>Transition rate to implementation and confirmation</td>
<td>GNP per capita Ethnic heterogeneity Size of old technology installed base Time of trial</td>
<td>Digital telecommunication switches</td>
<td>More than 160 countries</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Method/Model</td>
<td>Countries/Variables</td>
<td>Findings/Comments</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Kumar and Krishnan (2002)</td>
<td>Cumulative adopters, t and m in own country and cumulative adopters in other country (for cross-country effects)</td>
<td>CD player, cellular phone, microwave oven, home computer Belgium, Germany, Norway, Denmark, Finland, United Kingdom, France</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
<td></td>
</tr>
<tr>
<td>Talukdar et al. (2002)</td>
<td>p, q, and a (Bass model)</td>
<td>VCR, CD player, microwave, camcorder, fax machine, cellular phone Canada, Mexico, US, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, UK, Australia, China, Hong Kong, India, Malaysia, Philippines, Singapore, South Korea, Thailand, Argentina, Brazil, Chile</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
<td></td>
</tr>
<tr>
<td>Tellis et al. (2003)</td>
<td>Time to takeoff</td>
<td>CD player, color TV, computer, dishwasher, dryers, freezer, microwave oven, refrigerator, VCR, washing machine Denmark, Norway, Sweden, Finland, Ireland, Belgium, Switzerland, Austria, Netherlands, Germany, Italy, Spain, France, United Kingdom, Greece, Portugal</td>
<td>Sales of most new products display a distinct takeoff in various European countries, at an average of 6 years after introduction. The time to takeoff varies substantially across countries. It is almost half as long in Scandinavian countries as in Mediterranean countries. While culture partially explains inter-country differences in time to takeoff, economic factors are neither strong nor robust explanatory factors.</td>
<td></td>
</tr>
<tr>
<td>Van den Bulte and Stremersch (2004)</td>
<td>q/p</td>
<td>Individualism, uncertainty avoidance, power distance, masculinity, competing standards, income inequality 52 consumer durables (e.g., Color Television, VCR, cellular telephone, microwave oven) 28 countries</td>
<td>The q/p ratio is positively associated with income inequality of a country, supporting the heterogeneity-in-thresholds interpretation. It also varies with four of the Hofstede dimensions of national culture, supporting the social contagion interpretation. The study also finds that products with competing standards have a higher q/p ratio. Finally, it finds effects of national culture only for products without competing standards.</td>
<td></td>
</tr>
</tbody>
</table>

(1) The findings stated in the last column of this table for each author (team) are based on the statements of the authors and do not necessarily indicate that the authors of the present paper agree with the claims made by these prior papers. We also focus on their substantive contribution and not so much on their possible methodological contribution.

(2) The symbols p, q, m, a, and t* are the coefficient of innovation, imitation, market potential, penetration potential and time to peak sales, respectively, in the Bass diffusion model.

(3) We only included articles that appeared in major marketing journals, such as *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science*, and *International Journal of Research in Marketing*. 

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Method/Model</th>
<th>Countries/Variables</th>
<th>Findings/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar and Krishnan (2002)</td>
<td>Cumulative adopters, t and m in own country and cumulative adopters in other country (for cross-country effects)</td>
<td>CD player, cellular phone, microwave oven, home computer Belgium, Germany, Norway, Denmark, Finland, United Kingdom, France</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
</tr>
<tr>
<td>Talukdar et al. (2002)</td>
<td>p, q, and a (Bass model)</td>
<td>VCR, CD player, microwave, camcorder, fax machine, cellular phone Canada, Mexico, US, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, UK, Australia, China, Hong Kong, India, Malaysia, Philippines, Singapore, South Korea, Thailand, Argentina, Brazil, Chile</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
</tr>
<tr>
<td>Tellis et al. (2003)</td>
<td>Time to takeoff</td>
<td>CD player, color TV, computer, dishwasher, dryers, freezer, microwave oven, refrigerator, VCR, washing machine Denmark, Norway, Sweden, Finland, Ireland, Belgium, Switzerland, Austria, Netherlands, Germany, Italy, Spain, France, United Kingdom, Greece, Portugal</td>
<td>Sales of most new products display a distinct takeoff in various European countries, at an average of 6 years after introduction. The time to takeoff varies substantially across countries. It is almost half as long in Scandinavian countries as in Mediterranean countries. While culture partially explains inter-country differences in time to takeoff, economic factors are neither strong nor robust explanatory factors.</td>
</tr>
<tr>
<td>Van den Bulte and Stremersch (2004)</td>
<td>q/p</td>
<td>Individualism, uncertainty avoidance, power distance, masculinity, competing standards, income inequality 52 consumer durables (e.g., Color Television, VCR, cellular telephone, microwave oven) 28 countries</td>
<td>The q/p ratio is positively associated with income inequality of a country, supporting the heterogeneity-in-thresholds interpretation. It also varies with four of the Hofstede dimensions of national culture, supporting the social contagion interpretation. The study also finds that products with competing standards have a higher q/p ratio. Finally, it finds effects of national culture only for products without competing standards.</td>
</tr>
</tbody>
</table>

(1) The findings stated in the last column of this table for each author (team) are based on the statements of the authors and do not necessarily indicate that the authors of the present paper agree with the claims made by these prior papers. We also focus on their substantive contribution and not so much on their possible methodological contribution.

(2) The symbols p, q, m, a, and t* are the coefficient of innovation, imitation, market potential, penetration potential and time to peak sales, respectively, in the Bass diffusion model.

(3) We only included articles that appeared in major marketing journals, such as *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science*, and *International Journal of Research in Marketing*. 

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Method/Model</th>
<th>Countries/Variables</th>
<th>Findings/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar and Krishnan (2002)</td>
<td>Cumulative adopters, t and m in own country and cumulative adopters in other country (for cross-country effects)</td>
<td>CD player, cellular phone, microwave oven, home computer Belgium, Germany, Norway, Denmark, Finland, United Kingdom, France</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
</tr>
<tr>
<td>Talukdar et al. (2002)</td>
<td>p, q, and a (Bass model)</td>
<td>VCR, CD player, microwave, camcorder, fax machine, cellular phone Canada, Mexico, US, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, UK, Australia, China, Hong Kong, India, Malaysia, Philippines, Singapore, South Korea, Thailand, Argentina, Brazil, Chile</td>
<td>The authors found evidence of significant lead lag, lag lead, and simultaneous cross-country interaction effects. These cross-country interaction effects are affected by similarity (cultural and economic) between countries. Developing countries have a slower adoption rate, compared to that of developed countries. PPP, urbanization and international trade of a country affects a new product’s penetration potential. Information access and introduction lag affects the coefficient of external influence. Heterogeneity in ethnicity and introduction lag affects the coefficient of internal influence.</td>
</tr>
<tr>
<td>Tellis et al. (2003)</td>
<td>Time to takeoff</td>
<td>CD player, color TV, computer, dishwasher, dryers, freezer, microwave oven, refrigerator, VCR, washing machine Denmark, Norway, Sweden, Finland, Ireland, Belgium, Switzerland, Austria, Netherlands, Germany, Italy, Spain, France, United Kingdom, Greece, Portugal</td>
<td>Sales of most new products display a distinct takeoff in various European countries, at an average of 6 years after introduction. The time to takeoff varies substantially across countries. It is almost half as long in Scandinavian countries as in Mediterranean countries. While culture partially explains inter-country differences in time to takeoff, economic factors are neither strong nor robust explanatory factors.</td>
</tr>
<tr>
<td>Van den Bulte and Stremersch (2004)</td>
<td>q/p</td>
<td>Individualism, uncertainty avoidance, power distance, masculinity, competing standards, income inequality 52 consumer durables (e.g., Color Television, VCR, cellular telephone, microwave oven) 28 countries</td>
<td>The q/p ratio is positively associated with income inequality of a country, supporting the heterogeneity-in-thresholds interpretation. It also varies with four of the Hofstede dimensions of national culture, supporting the social contagion interpretation. The study also finds that products with competing standards have a higher q/p ratio. Finally, it finds effects of national culture only for products without competing standards.</td>
</tr>
</tbody>
</table>
called the takeoff at its start and by slowdown at its termination (Golder & Tellis, 1997; Golder & Tellis, 2004).

In the context of increasing globalization, the challenge facing managers is how to sustain that growth across countries with dramatically varying demand. The strategy depends on answers to the following questions.

1. Is growth of new products similar or substantially different across countries?
2. If different, does economics or culture influence the pattern of growth across countries?
3. What are the implications of the answers to these questions for:
   - the choice of a waterfall (introducing in different countries at different times) versus sprinkler (introducing in all countries at the same time) strategy for new products;
   - global versus local marketing of a new product;
   - managing a firm’s expectations about new product growth?

The current study of the sales growth of 10 consumer durables in 16 European countries aims to answer these questions. It advances the literature on international diffusion of new products that Table 1 summarizes.1

Most of the articles in this tradition use parameters from the Bass model to study variation across countries. The Bass modeling tradition treats diffusion as an outcome of external \( p \) and internal \( q \) influences. These two parameters can then be combined to estimate the speed of diffusion. In their study of the international takeoff of new products, Tellis, Stremersch, and Yin (2003) use a different approach, grounded in affordability theory (Golder & Tellis, 1997; Golder & Tellis, 2004). The underlying premise of the theory is that the changing affordability of a new product, as its price declines over time, determines the speed and growth in its sales (Tellis & Golder, 2001). The key measure that they use is the time to takeoff in sales of the new product. However, takeoff is followed by strong growth in new product sales. None of the prior studies has addressed the international pattern and drivers of duration and rate of the growth in new product sales that follows takeoff. This is an important void for several reasons.

First, cross-country variation in growth may differ from cross-country variation in time to takeoff. Second, would culture or economics be the primary explanation for such cross-country variations in growth? Third, if economics rather than culture explains these differences, then how would one reconcile that with culture being the primary driver of inter-country differences in time to takeoff?

The current paper makes three contributions to the literature. First, it proposes two new, direct, and fruitful operationalizations of growth: duration of growth and rate of growth during the growth stage of the product life cycle. Second, it reconciles the alternate explanations (culture and economics) for inter-country differences in growth and time to takeoff. Third, it provides implications on how to manage new product growth across countries.

This paper is organized as follows. The next section presents the theoretical background and our research hypotheses. The third section discusses our data. The fourth section presents the empirical results. The fifth section discusses our findings, and considers the study’s limitations and implications for marketing management and future research.

2. Why growth varies: theory and hypotheses

This section explores the reasons why the growth of new products may vary across countries. Our focal criterion variable throughout is growth. However, we can measure the growth of new products by two indices: (1) the average rate of growth during the growth stage; and (2) the duration of the growth stage.

The two indices of growth may be related to each other if market penetration is held constant. A high growth rate will imply a short growth stage and vice
versa. \footnote{Such a strict relationship is only valid when ultimate penetration levels at maturity do not differ across countries. When this is not the case, it can be that growth duration is short and growth rates are low, as the category fizzes out and fails to achieve enough penetration in a country. In our sample of countries, we find no significant differences across countries in ultimate penetration level that are consistent across categories. We also find empirically that growth duration and growth rate are negatively related. The median and average correlation between the two measures across product categories within countries is −0.46 and −0.22, respectively. The median and average correlation between the two measures across countries within product categories is −0.44 and −0.37, respectively. Thus, for the sake of brevity, we will treat growth duration and growth rate as negatively related to each other, and the empirical analysis shows this to be a valid perspective for our sample of countries.} In the interest of parsimony, we will discuss the theory and hypotheses with respect to the growth rate only. Because of the negative relationship between the two indices, we expect the logic and hypotheses to reverse for the duration of the growth stage. Nevertheless, since the two indices are not necessarily equivalent, our empirical analysis will explore the effects of the causal variables on each index. This exercise increases the validity of the tests and the reliability of our conclusions.

To explain variation in growth rates across countries, we include two sets of predictors, (1) economics and (2) culture of the country. We discuss each in turn.

2.1. Economics

Economic theory suggests that two factors may be pertinent to how new products grow across countries: economic wealth and income inequality. We next explore the effects of these two factors on the growth rate of new products across countries.

2.1.1. Economic wealth

Economic wealth refers to the average wealth of the people of a country. Wealth determines to what extent the population at large can afford to buy new products. Indeed, a general finding in adoption research is that high-income consumers are generally the first to adopt a new product (Rogers, 1995). The reason is that wealthier consumers can better afford a new product than poorer people can, especially early in its life cycle when it is still priced highly. Wealthier consumers can also better afford the risk of adopting a new product earlier than poorer consumers (Dickerson & Gentry, 1983). In addition, wealthier countries often have better media infrastructures. Consequently, consumers can be more easily informed (Beal & Rogers, 1960) and convinced (Katz & Lazarsfeld, 1955) of the benefits of the new product. Also consumers may learn of the adoption and satisfaction of other consumers more rapidly. Therefore, we expect higher growth in wealthy countries, as compared to poor countries. Thus, we hypothesize:

**H1.** New products grow faster in wealthy countries than in poor countries.

2.1.2. Income inequality

In addition to the average wealth of a population, the distribution of wealth or income may also affect the growth of new products. Even if a country is wealthy, uneven income distribution may imply that many segments fall below the threshold to buy a new product. As a result, in such countries, new products may remain unaffordable for large parts of the population causing sluggish growth. So, we hypothesize:

**H2.** New products grow slower in countries with high income inequality than in countries with low income inequality.

2.2. Culture

Culture is the collective programming of the mind that distinguishes the members of one human group from another. Although we can, in principle, use the word culture for any social group, here, we reserve it for societies or countries. Prior research suggests that distinct cultural traits underlie systematic differences in consumer behavior (de Mooij & Hofstede, 2002; Lynn, Zinkhan, & Harris, 1993; Steenkamp, 2001), especially differences in response to new products (Jain & Maesincee, 1998; Steenkamp, ter Hofstede, & Wedel, 1999; Yeniyurt & Townsend, 2003). We focus on three dimensions of culture that have been found to be relevant to potential differences in new product growth, namely uncertainty avoidance, masculinity and religion.
2.2.1. Uncertainty avoidance

Uncertainty avoidance refers to the extent to which the members of a culture feel threatened by uncertain or unknown situations (Hofstede, 2001). In uncertainty avoidant cultures, risk taking is limited to known risks (of which the probability is known), while in cultures low in uncertainty avoidance, risk taking includes unknown risks (of which the probability is not known). Uncertainty avoidant cultures are extremely conservative, in which people generally resist change (Hofstede, 1980). As new products involve change not only in the material realm but also in people’s attitudes and behaviors, one can expect cultures high on uncertainty avoidance to show low innovativeness and thus slower growth. Therefore, we hypothesize:

\[ \text{H3a. New products grow slower in countries high in uncertainty avoidance than in countries low in uncertainty avoidance.} \]

However, uncertainty avoidance may not only affect intrinsic innovativeness of a culture, but it may also affect the extent to which it is important for members of a culture to learn from one another. Non-adopters can learn of a new product’s features by observing other people’s adoption of a new product and interacting with them. This behavior reduces non-adopters’ uncertainty and triggers their adoption of the new product. Such uncertainty reduction is more important for uncertainty avoidant cultures than for cultures low in uncertainty avoidance. Therefore, one would expect that members of the former cultures are more influenced by prior adopters than members of the latter culture (Van den Bulte & Stremersch, 2004). This leads to a faster diffusion and thus faster growth of the new product in an uncertainty avoidant country as compared to a country low in uncertainty avoidance. This leads to an alternative hypothesis:

\[ \text{H3b. New products grow faster in countries high in uncertainty avoidance than in countries low in uncertainty avoidance.} \]

2.2.2. Masculinity

Masculinity refers to the sex role pattern in social groups whether it is characterized by male (e.g., assertive) or female (e.g., nurturing) attributes. Masculinity is a value system shared especially by the majority of the people in the middle class of a society (Hofstede, 1980). In masculine societies, people are more materialistic and admire successful achievers (Hofstede, 1983). In such societies, consumers may autonomously adopt new products faster, since it allows them to show off achievement, and thus these countries experience faster new product growth. Also, display of status is more important in masculine societies and as new products may be accepted out of status considerations (Van den Bulte & Stremersch, 2004), masculine societies may adopt new products faster and thus again show faster growth. Therefore, we hypothesize:

\[ \text{H4. New products grow faster in masculine countries than in feminine countries.} \]

2.2.3. Religion

The religion of a society is a cultural trait that may have substantial effects on the growth of new products in a country (e.g., Tellis et al., 2003). In the Western European countries—the context of the present study—the main religious faiths are Catholicism and Protestantism. There is evidence in sociology that Protestant religions are more supportive of a high need for achievement than is the Catholic faith (McClelland, 1961; Weber, 1958). A high need for achievement makes people value effectiveness and efficiency highly. New consumer durables make work in the home more efficient and effective. Thus, a higher need for achievement will encourage people to adopt new consumer durables faster. Thus, we hypothesize:

\[ \text{H5. New products grow faster in countries with a larger proportion of Protestants than those with a smaller proportion of Protestants.} \]

2.3. Other variables

Though they are not of main interest to us, we also control for three other variables. A first control variable is the product class, whether the product is a brown or white good. Brown goods are electronic goods such as TVs and digital cameras, and white goods are kitchen and laundry appliances. Brown goods typically are more glamorous and appealing than white goods because they are more visible,
enjoyed by all members of the household, and more frequently discussed in social circles, than are white goods. So, we expect that brown goods will have higher growth rates than white goods. A second control variable is lagged market penetration. We expect that as products reach a higher market penetration, they grow more slowly.

A third control variable is the lag with which the product is introduced in a country. We expect that the later a product is introduced in a country—compared to the lead country—the faster it will grow relative to countries with early introduction. This expectation can be supported through several arguments. First, manufacturing and marketing expenses fall at a constant rate the more experience suppliers have accumulated. In the presence of competition, typical of most consumer products, prices tend to fall at a similar constant rate. A large number of studies support this thesis (for a recent overview, see Argote, 1999). Therefore, lags in introduction of a new product in a particular country can be seen as an advantage for growth of that product in that country. However, differences in launch time may also capture other effects, such as changes in purchasing power and household formation rate (Van den Bulte, 2000), knowledge dissemination through reverse engineering or cross-country influences, among others.

3. Data

This section describes our data collection and measures.

3.1. Data collection

This study uses the database of historical data on sales of new consumer durables from Tellis et al. (2003). This database—composed from sources, such as Euromonitor, GfK, The Economist Intelligence Unit, Tablebase, archives and publications of associations of appliance manufacturers and William P. Putsis, Jr.—contains sales data on 10 consumer durables (refrigerator, washing machine, freezer, dishwasher, color TV, dryer, VCR, computer, CD player, and microwave oven) across 16 European countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the UK). For our purposes, we had complete data on 114 country–category pairs. Our database covers annual data from the period 1950–2000.

Our key sources of data for the independent variables are the Statistical Yearbook of the United Nations, the Penn World Table, the World Bank Statistics, Eurostat Review, and individual sources, such as Parker (1997) and Hofstede (1980, 2001).

3.2. Measures

This subsection explains the measures for the dependent and independent variables in our model.

3.2.1. Dependent variables

As stated at the beginning of the hypotheses section, our hypotheses relate to two different dependent variables, namely duration of the growth stage (how long does growth last?) and the growth rate during the growth stage (at which rate do sales grow?).

3.2.1.1. Duration of growth stage. The duration of the growth stage of the product life cycle is the time that elapses between takeoff and the end of the growth stage. Takeoff is the start of the growth stage of the life cycle characterized by a rapid growth in sales. To measure takeoff, we adopted the threshold rule developed by Tellis et al. (2003). They define the threshold for takeoff through a standard plot of growth in sales for various levels of market penetration. They operationalize takeoff as the first year a product’s growth in sales crosses the threshold. The end of the growth stage is one period before sales slow down (decline). To determine the location of the end of the growth stage, we adopt a rule developed by Golder and Tellis (2004). By this rule, the end of growth is the first year, after takeoff, after which two consecutive years occur with lower sales. To show our measure of the duration of the growth stage, Fig. 1 graphs the sales evolution of microwaves in Germany and the UK from introduction to 1990. We have arrows in a full line to indicate the start of the growth stage (takeoff) and arrows in dotted line to indicate the end of the growth stage (one period before slowdown).
3.2.1.2. Growth rate during growth stage. With growth rate during the growth stage, we refer to the average growth rate over the growth stage (excluding the year of takeoff). In symbols:

\[
GR_{ij} = \frac{1}{T} \sum_{t=1}^{T} \frac{S_{ijt} - S_{ijt-1}}{S_{ijt-1}}
\]

with \( GR_{ij} \) representing growth rate of category \( i \) in country \( j \), \( T \) the number of periods in the growth stage and \( S_{ijt} \) unit sales of category \( i \) in country \( j \) at time \( t \). This measure of growth rate is independent of model assumptions and more intuitive than the measure developed by Van den Bulte (2000). Ideally, a time-varying growth measure would preserve the information in the data. However, we use an average growth measure for three reasons. First, growth rates are highly volatile over time. As such, fitting any model to the data with time varying growth rates is quite complex and cumbersome. Second, our focus is inter-country differences in growth rate, not variation in growth rates across time. Third, an average growth measure has intuitive appeal and is easy to interpret.

3.2.2. Independent variables

For the cultural variables of uncertainty avoidance and masculinity, we used Hofstede’s (1980) measures, as these match the time period covered by our data (post 1950) and provide measures on all countries on which we have data. Readers may refer to the original work of Hofstede (1980) or its most recent edition (Hofstede, 2001). For the cultural variable of Protestantism, we used the percentage of Protestants as provided by Parker (1997).

We measured economic wealth by GDP per capita in thousands of US dollars. We also included real GDP per capita in constant dollars, adjusted for changes in the terms of trade (we used the 1985 international prices for domestic absorption and current prices for exports and imports). This measure gave similar results. We measured income inequality
by the GINI Index, as extracted from the World Bank database (Deininger & Squire, 1996). To maximize consistency across countries, we selected the GINI coefficient based on net income, number of households, and national coverage.

We also included several other variables. First, we account for differences between brown and white goods and included the product class as a dummy variable, coded 1 for white goods and 0 for brown goods. Second, for lagged market penetration, we used the lagged average possession of the product by households in the country. Our sources (GfK and Euromonitor) provided us the market penetration for the white goods. For brown goods, we calculated the market penetration as follows:

\[
\text{penetration}_t = \text{penetration}_{t-1} + \left( \frac{\text{sales}_t - \text{sales}_{t-r}}{\text{number of households}_t} \right),
\]

where \(r\) is the estimated average repurchase time for a product in a particular category. The above measure for penetration is a rough proxy as (1) it does not adjust for repeat purchase and may thus overestimate penetration; and (2) the average repurchase time was estimated by us, based on our own judgement, and kept constant over time.\(^3\) Third, we control for the introduction lag, which is the lag with which a new product was introduced in a country as compared to the country it was first introduced in Europe in.\(^4\)

Since the variables in our model include both time-varying and time-invariant variables, we need to point out clearly which variables are of which type.

- **Time-varying variables are:** lagged market penetration, economic wealth, income inequality.

---

\(^3\) As average repurchase times during the growth stage, we used 4 years for personal computer, 5 years for CD-Player, 6 years for VCR, and 8 years for color television. Note that these estimates may appear high, as compared to present repurchase rates (in the maturity stage), but credible for repurchase times relatively early on in the product life cycle. Also, our results were not sensitive to changes in these repurchase times.

\(^4\) We also checked empirically whether the lag time versus the original US introduction—for all our products the US was the first country in which the product was introduced in (either alone or simultaneously with other countries)—was an explanation for growth rate and duration (as it may also capture economies of experience). We found that this lag time had no effect and for sake of brevity do not report on it in detail.
Time-invariant variables are: product class, introduction lag, uncertainty avoidance, masculinity, Protestantism.

Table 2 presents means of variables by country and Table 3 presents overall means of and correlations between variables.

4. Pattern and drivers of sales growth: empirical results

Recall that the first two goals of this study were to examine the pattern of sales growth across Europe as to: (1) whether there are country-specific differences in the duration and speed of growth across European countries; and (2) if yes, what factors explain these inter-country differences. We first discuss the descriptive statistics that aim to answer the first question. Then we continue with our explanatory analyses on the drivers of inter-country differences.

4.1. Describing the sales growth pattern across Europe

We describe the pattern of sales growth across Europe using the two dependent variables we identified above, duration of growth stage and growth rate. Before we discuss each, we caution the reader that these descriptive analyses are exploratory because they do not control for other influences, such as product category and introduction timing. Also, the standard deviations are fairly large, which is a caveat against strong conclusions, especially on the comparison of individual countries.

4.1.1. Duration of growth stage

We examine the average duration of the growth stage across countries in Table 4a. The duration of the growth stage is the time that elapses between takeoff and slowdown. From this analysis, we may conclude that Nordic countries have relatively shorter growth stages, as compared to other countries (see Table 4a). The average growth stage in Nordic countries (Denmark, Norway, Sweden, and Finland) is about 8 years, which is lower than the average across all countries of about 10 years. Mid-European and Mediterranean countries have comparable durations, since the growth stage lasts little over 10 years for both these country groups.

4.1.2. Growth rate during growth stage

A second interesting metric with which we can examine the pattern of sales growth across Europe is the rate at which sales grow during the growth stage. We calculate the average growth rate of the new products in our sample for each country. To conduct this analysis, we excluded the year of takeoff itself, as growth rates in the takeoff year may be very large and thus may dominate the average. The country ranking
is fairly robust to this exclusion, thus the ranking is similar when the takeoff year is not excluded. The results are in Table 4b.

From Table 4b, we can see that the average growth rate in the growth stage of the product life cycle across all countries is equal to 41.1%. Nordic countries (45.9%) generally have the steepest growth during the growth stage, followed by Mid-European (41.4%) and then Mediterranean (36.0%) countries.

4.2. Drivers of the sales growth pattern across countries

We hypothesized on theoretical drivers of the duration of the growth stage and on the rate at which sales grow during the growth stage. We first discuss the models we employ for growth duration and rate after which we turn to the results we obtained.

4.2.1. Models

The most appropriate model for growth duration is a hazard model. Since some of the predictors, as we explained in the Data, are time-varying, we use a parametric hazard model\(^5\) and not a proportional hazard model (Cox, 1972; Jain & Vilcassim, 1991). The duration of growth is modeled through a Weibull specification (monotonic hazard). A technical appendix on the full model specifications is available from the authors upon request. Note that we also estimated a Weibull specification with Gamma heterogeneity to capture unobserved heterogeneity in our estimates. Results from these estimations were very similar to a regular Weibull model. Thus, we opted for the more parsimonious specification without Gamma mixing. We also examined the robustness of our assumption of a Weibull distribution by estimating other hazard models with other baseline distributions (such as the Log-logistic and Gamma) and found our results to be robust to the choice of the baseline.

To model the growth rate across countries, we use a traditional linear regression model, with the same set

---

\(^5\) Note that our model (also see the appendix) is a continuous time hazard model, while our data are strictly speaking discrete time data. However, prior research has shown that “…the discrete-time method…will virtually always give results that are quite similar to the continuous time methods.” (Allison, 1984, p. 22).
of independent variables as in the hazard model for growth duration.6

To illustrate robustness, we specify four different models for both growth duration and growth rate (see Table 5). The first model only includes the control variables, product class, lagged penetration and introduction lag. The second model includes the control variables and the economic variables. The third model includes the control variables and the culture variables. The fourth model includes all variables.

4.2.2. Results

The results are in Table 5. Note that for the hazard model, positive \( \beta \) coefficients increase duration of growth and negative \( \beta \) coefficients decrease duration of growth, while for the OLS model, positive \( \beta \) coefficients increase growth rate and negative \( \beta \) coefficients decrease growth rate. For each model specification, we provide the parameter estimates with standard errors in-between brackets, the number of observations included, and fit statistics (LL and AIC for the hazard model; \( R^2 \)-squared and adjusted \( R^2 \)-squared for the OLS model). For the OLS model, we report standardized coefficients. For the hazard model, Table 5 reports unstandardized coefficients, while in the text in-between brackets, we also report the change in the hazard ratio—denoted as \( \Delta \)—associated with each independent variable. The change in hazard ratio represents the percentage change in the hazard ratio given a one-unit change in the independent variable. This value is equal to \( 100 \times (e^{\beta} - 1) \) (see Tellis et al., 2003). We next discuss our findings per set of variables.

The first hazard and OLS model only includes the control variables, product class, lagged penetration, and introduction lag. We find that lagged penetration has no influence, introduction lag has a limited influence, which is not robust to model specification, while white goods have longer growth stages than brown goods (\( \beta_{\text{CLASS}}=0.46 \) with \( p<0.01 \); \( \Delta_{\text{CLASS}}=-36.9 \)) and grow at a slower rate (\( \beta_{\text{CLASS}}=-0.27 \)). Therefore, we can conclude that brown goods grow faster and have shorter growth durations than white goods.

The second hazard and OLS models show that, as hypothesized in H1, economic wealth affects both growth duration (\( \beta_{\text{GDP}}=-0.03 \); with \( p<0.01 \); \( \Delta_{\text{GDP}}=3.00 \)) and growth rate (\( \beta_{\text{GDP}}=0.53 \); with \( p<0.01 \)). Income inequality affects neither growth rate nor duration. Thus, we conclude that consistent with H1, economic wealth negatively affects growth duration and positively affects growth rate, while income inequality does not have a significant effect on growth duration nor growth rate, in contrast to H2.

The third hazard and OLS models show that culture does not have an effect on growth duration and growth rate. Therefore, we conclude that culture consistently does not have an effect on growth duration and growth rate, in contrast to H3a, H3b, H4, H5.

The fourth hazard and OLS models include all variables. These full models mostly confirm our findings from the nested models. Note that none of the full or nested models suffers from harmful multicollinearity, as the highest condition index is equal to 3.5, which is much lower than the threshold value of 30, recommended by Belsley, Kuh, and Welsh (1980).

All these results allow us to formulate one overall conclusion: Economic wealth has a strong and negative effect on growth duration and a strong and positive effect on growth rate, while culture and income inequality do not play a significant role.

4.2.3. Further analyses

We find strong support for our expectations regarding economic wealth. However, our theoretical prediction was based on two arguments: (1) affordability; and (2) availability of media infrastructure. To examine which of the two drives our results, we estimated an additional model that included a media infrastructure variable, which was an index of the number of TVs, radios and newspapers in a country. When this variable was included in the model, we found that it was significant in the regression analysis, but insignificant in the hazard analysis. In both models, its inclusion lowered the significance of the effect of economic wealth slightly, while it did remain significant. This analysis provides some (albeit incomplete) evidence that both theoretical mecha-

---

6 For this analysis, we averaged the time-varying independent variables over the growth stage.
Second, we only included two of the four Hofstede dimensions of national culture. To examine post hoc whether this choice has affected our conclusions, we estimated a model that included all four Hofstede dimensions (in addition to the other variables in our full model 4). In this model, all four cultural dimensions were insignificant, while economic wealth remained strongly significant ($p < 0.01$). To check if this finding is an artifact of the Hofstede framework, we ran all models with the cultural dimensions of Schwartz instead of those of Hofstede (see Schwartz, 1994). We find that none of the Schwartz dimensions significantly affect either the duration or rate of growth, while the effect of economic wealth again is strongly significant.

5. Discussion

5.1. Findings

Our findings have answered two of the research questions posed at the outset of this paper. First, we questioned if the pattern of the growth of new products differed across countries? We found that there are strong differences across countries in both
growth rate and growth duration. This findings has never been reported before. It complements past findings about strong differences across European countries in the Bass diffusion parameters (Gatignon, Eliashberg, & Robertson, 1989) and time to takeoff (Tellis et al., 2003). However, we also found that differences among geographic regions—Nordic (Sweden, Denmark, Norway, and Finland), Mid-European and Mediterranean—are relatively small, especially in growth duration. This finding is also new and complements past research that has found very strong differences across these regions in time to takeoff (Tellis et al., 2003).

Second, we questioned if economics or culture explained the differences in the pattern of growth across countries. We found that economic wealth primarily explains the inter-country pattern of growth. Culture does so to a far lesser extent than economic wealth. This is exactly the opposite of the findings of Tellis et al. (2003) who found that culture explains time to takeoff across countries better than economic wealth. We theorize that the reason for these contradicting results is that takeoff is a phenomenon very early in the product life cycle, typically below 2–3% market penetration. On the other hand, growth is later in the product life cycle, somewhere between 3% and 35% market penetration (Mahajan et al., 1990). In the classical adoption terminology (Rogers, 1995), innovative consumers that adopt before takeoff may be especially driven by cultural factors, while early adopters and early majority may be more driven by affordability concerns. Therefore, international takeoff patterns may be predominantly driven by cultural traits of countries, while international growth patterns may be predominantly driven by the economic wealth of countries. Indeed, our results seem to complement those of Talukdar, Sudhir, and Ainslie (2002), who also found a strong effect of economics on diffusion patterns. Thus, our explanation helps to reconcile contradictions in prior work in this area.

5.2. Managerial implications of findings

At the outset of this paper, we also formulated three research questions relating to the management of new product growth: (1) the choice of a waterfall versus sprinkler strategy for new product introduction; (2) the global versus local marketing of a new product; and (3) managing expectations on new product growth. We discuss each in turn.

5.2.1. Choice of waterfall versus sprinkler strategy

A sprinkler strategy is one in which a firm introduces in all countries at the same time. A waterfall strategy is one in which a firm introduces in different countries at different times. The rationale for each is the following:

• There are two advantages for a sprinkler strategy. First, a sprinkler strategy can maximize revenues by fully exploiting economies of scale and experience in R&D and manufacturing. It does so by exposing the new product to a maximum number of markets as soon as possible, thus tapping the widest possible scale of operation from the outset. Second, if competition is a threat, then a sprinkler strategy may pre-empt competitive moves in at least some countries, thus maximizing share of market.

• There are two key advantages of a waterfall strategy. First, launching a new product requires investments in manufacturing, inventory, advertising, distribution, sales force, and staff. A waterfall strategy requires a much lower investment than a sprinkler strategy, because the new product is introduced in only a subset of countries. If the product fails in those countries, a manager need not launch in the remaining countries, thus surely saving the investment in the latter countries. Second, because revenues and profits from an early market can be used for investment in a subsequent market, a waterfall strategy also greatly lowers the pressure on cash flow relative to a sprinkler strategy. Now, for any new product, the outcome is uncertain, both in terms of annual sales and ultimate success. Therefore, the lower startup investment and the lower pressure on cash flow translates into lower risk in a waterfall strategy than in a sprinkler strategy. In Europe, one can think of two possible waterfall strategies, one (the Current Waterfall, as that is what companies currently do) of introducing in the large countries first (see Putsis, Balasubramanian, Kaplan, & Sen, 1997) and one (the North-to-South Waterfall) of introducing first in the Nordic countries, then in the Mid-European countries and finally in the Mediterranean countries (see Tellis et al., 2003).
Thus, the essential tradeoff between these introduction strategies boils down to one between maximizing revenues and minimizing risk. The literature (Kalish, Mahajan, & Muller, 1995; Putsis et al., 1997; Tellis et al., 2003) is unclear about which strategy is optimal. Through a simulation that uses the results of the present study and those of Tellis et al. (2003), one can predict what the levels of sales and risk would be for each year from introduction, for any possible introduction strategy. Details from this simulation are available from the authors as a technical note. Here, we only briefly outline the logic and the intuition of a few results (applied to the Freezer category).

The first part of Fig. 2 shows the evolution of the sales level. As our arguments above indicate, the market size effect of a sprinkler strategy clearly dominates the positive—but small—cross-country learning effect and thus generates more sales. The second part of Fig. 2 shows a risk index for companies under the three strategies. We define this risk index as the product of investments in manufacturing and the standardized variance in sales for each year from introduction. We compute the variance in sales for each year from introduction, as the variance in sales for all similar categories, at that year, in all countries in which the target category would be introduced. Fig. 2 shows that, while the sprinkler strategy generates more sales, it also incurs more risk than the waterfall strategy. A North-to-South waterfall involves the least risk. Analysis of the results suggests two reasons for this low risk. First, investments are limited to a constrained set of small (Nordic) countries, involving smaller investments and lower variance. Second, the expansion to other (larger and higher variance) countries is spread over a long period of time.

This framework shows that the tradeoff between a waterfall and sprinkler strategy reduces to a tradeoff between sales maximization and risk minimization. In our consultations with researchers and managers, we find that researchers tend to favor a rapid deployment across all countries to maximize sales and market share. However, managers are deeply concerned about the risk of failure. They have no certainty of the success of their new products, especially early on. Even if they are convinced that the new product will succeed, they remain uncertain of the dates of takeoff and the rate of growth.

5.2.2. Global versus local marketing of a new product

Our results show that there are dramatic differences across countries in the growth pattern of new products. This is a strong argument in favor of localized marketing strategies. It seems obvious that when countries are in different stages of the product life cycle (introduction–growth–maturity) and experience different growth rates, they need a different market approach. For instance, in the introduction stage, investments may be rather limited and targeted towards informing consumers of the new product and entice innovators to try it out. However, in the growth stage, firms have to gear up for a larger market that is looming and have to target the mass market. Global marketing strategies would ignore such inter-country differences and thus may be suboptimal. The least we would expect global companies to do—should they wish to maintain standardized marketing strategies across the globe—is to adjust the actual calendar time in which the strategy is deployed to the stage of the life cycle the new product is in a specific country.

5.2.3. Managing expectations on new product growth

The many descriptive statistics we offer in this paper also allow managers to set their expectations at a more realistic level. From our own experiences, often, managers underestimate the time it will take for a product to take off, after which they overestimate the time at which sales will start to slow down. This paper gives managers in consumer electronics and household appliances sound expectations as to what sales pattern to expect of new consumer durables. Managers that have more realistic expectations can be expected to make better decisions.

---

7 Our risk index, in symbols is: Risk\(_{kt}\) = inv\(_{kt}\) × var(S\(_{ijkt}\))/mean(S\(_{ijkt}\)); in which Risk\(_{kt}\) represents the risk in time period \(t\) (\(t=1,\ldots,T\)) in scenario \(k\) ("sprinkler", "current waterfall", "North-to-South waterfall"); inv\(_{kt}\) represents the investments in time period \(t\) and in scenario \(k\); var(S\(_{ijkt}\)) and mean(S\(_{ijkt}\)) represent the variance and mean, respectively, of sales across other categories \(i\) and countries \(j\), under scenario \(k\), in time period \(t\). Our risk index thus accounts both for the magnitude of investments and for the variability of sales. While the former accounts for the total cost in the event of unused capacity (e.g., because of withdrawal of the new product), the latter accounts for the probability of this event occurring.
Fig. 2. (A) Unit sales comparison across the scenarios: freezers. (B) Risk comparison across the scenarios: freezers.
5.3. Limitations and future research

This study has several limitations. First, we studied growth in only the growth stage of the product life cycle. It may be fruitful to explore if the patterns and drivers of growth differ across all stages of the product life cycle (e.g., the saddle as in Goldenberg, Libai, & Muller, 2002). Second, we have some measures that are limited, while we do not have measures for some other important variables, such as regulation. Third, we separately estimated both models (growth duration and growth rate), although one may assume that they are interdependent. Fourth, we do not account for differences across countries in ultimate market penetration levels. We also only included successful products—products that in the end got adopted by the mass market—in our analysis. Fifth, we did not explore how the concepts in this paper—takeoff, growth, and slowdown—can be related to the Bass diffusion model parameters. Many of these limitations may suggest fruitful avenues for future research.

Acknowledgements

We greatly benefited from the comments of Jan-Benedict Steenkamp and Geert Hofstede. We also appreciated the many constructive comments of participants of the MSI and IJRM conference on Global Marketing in Noordwijk (The Netherlands), seminar participants at Erasmus University and Maastricht University, three anonymous IJRM reviewers and both special issue editors, Marnik Dekimpe and Don Lehmann.

References


A web of glass spans the globe. Through it, brief sparks of light incessantly fly, linking machines chip to chip, and people face to face (Cerf, 1991, p. 72)

1. Introduction

Marketers have become more and more interested in learning about, organizing, and managing virtual communities on their internet venues (Bagozzi & Dholakia, 2002; Balasubramanian & Mahajan, 2001). Such an interest stems not only from their ability to influence members’ choices, and to rapidly disseminate knowledge and perceptions regarding new products (e.g., Dholakia & Bagozzi, 2001), but also from the numerous opportunities to engage, collaborate with, and advance customer relationships actively in such forums. In the current research, consistent with the prevailing view (e.g., Rheingold, 2002; Wellman & Gulia, 1999), virtual communities are viewed as consumer groups of varying sizes that meet and

A social influence model of consumer participation in network- and small-group-based virtual communities

Utpal M. Dholakiaa,⁎, Richard P. Bagozziab, Lisa Klein Pearob

aRice University, Jesse H. Jones Graduate School of Management, 6100 Main Street, 314 Herring Hall-MS 531, Houston, TX 77005, USA
bCornell University, Cornell School of Hotel Administration, Ithaca, NY 14853, USA

Received 8 May 2003; received in revised form 1 September 2003; accepted 5 December 2003

Abstract

We investigate two key group-level determinants of virtual community participation—group norms and social identity—and consider their motivational antecedents and mediators.

We also introduce a marketing-relevant typology to conceptualize virtual communities, based on the distinction between network-based and small-group-based virtual communities. Our survey-based study, which was conducted across a broad range of virtual communities, supports the proposed model and finds further that virtual community type moderates consumers’ reasons for participating, as well as the strengths of their impact on group norms and social identity. We conclude with a consideration of managerial and research implications of the findings.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Virtual communities; Internet marketing; Consumer behavior; Electronic commerce; We-intentions

* Corresponding author. Tel.: +1 713 348 5376; fax: +1 713 348 6331.
E-mail addresses: dholakia@rice.edu (U.M. Dholakia),
bagozzi@rice.edu (R.P. Bagozzi), lkp22@cornell.edu (L.K. Pearo).

0167-8116/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
interact online for the sake of achieving personal as well as shared goals of their members.

Researchers have employed various theories such as social network analysis (e.g., Wellman & Gulia, 1999), life cycle models (e.g., Alon, Brunel, & Schneier Siegal, 2004), and motivational theories (e.g., Bagozzi & Dholakia, 2002) for studying virtual communities, examining such issues of marketing relevance as what draws participants to such communities, what they are used for, and how they influence the subsequent knowledge, opinions, and behaviors of participants. A common theme underlying many of these investigations is to better understand the nature and role of the social influence exerted by the community on its members (Alon et al., 2004; Postmes, Spears, & Lea, 2000; see Dholakia & Bagozzi, 2004 for a review).

Bagozzi and Dholakia’s (2002, hereafter B&D) study provides a useful starting point for framing our discussion since it adopted a marketing lens to identify two key social influence variables, group norms, and social identity, impacting virtual community participation. Using the social psychological model of goal-directed behavior (e.g., Perugini & Bagozzi, 2001) and social identity theory (e.g., Tajfel, 1978) as underlying frameworks, B&D conceptualized participation in virtual chat rooms as intentional social action involving the group. They modeled participants’ ‘we-intentions,’ i.e., intentions to participate together as a group, to be a function of individual (i.e., attitudes, perceived behavioral control, positive, and negative anticipated emotions) and social determinants (i.e., subjective norms, group norms, and social identity).

Despite the insights derived from their theorizing and empirical analysis, the following two limitations of the B&D framework are noteworthy and provide the motivation for the present research. First, B&D viewed the social influence variables to be exogenous constructs in their framework, i.e., they did not consider the antecedents of either group norms or social identity, two important predictors in their model. Understanding the antecedents of social influence is important since it is likely to provide significant managerial guidance regarding how to make virtual communities useful and influential for their participants. Second, B&D’s empirical study was limited to virtual chat rooms and did not consider or elaborate on the distinctions between different types of virtual communities or their implications for marketers. Indeed, marketers have narrowly conceived of virtual communities as commercially sponsored bulletin-boards or chat rooms on company websites (e.g., Thorbjørnsen, Supphellen, Nysveen, & Pedersen, 2002; Williams & Cothrel, 2000; cf. Catterall & Maclaran, 2001). Addressing these limitations, our objectives in the present research are three-fold.

First, building upon the B&D (2002) framework, we develop a social influence model of consumer participation in virtual communities. Like B&D (2002), the central constructs in our model are group norms and social identity, but unlike B&D, we not only consider the antecedents of social influence, but also include such mediating constructs as mutual agreement and accommodation among group members. We draw upon existing communication research regarding the motivational drivers of media use (e.g., Flanagin & Metzger, 2001; McQuail, 1987), philosophical writings on group action (Bratman, 1997; Tuomela, 1995) and social psychological research on social identity (e.g., Ellemers, Kortekaas, & Ouwerkerk, 1999; Tajfel, 1978) to develop our theoretical model.

Second, we present a marketing-relevant typology to conceptualize virtual communities within a firm’s internet venues that makes and elaborates on the distinction between network- and small-group-based virtual communities. In doing so, we also make the conceptual distinction between the venue where the virtual community meets, and the networks or small groups of individuals constituting the community. In our survey-based study conducted across a broad range of virtual communities, our proposed model is supported. We also find virtual community type—network- or small-group-based—to be a moderator, influencing both, the reasons why members participate, and the strengths of their impacts on group norms and social identity.

Finally, we consider the implications of our framework and the distinction made between network- and small-group-based virtual communities, for marketing practice. We elaborate on some of the trade-offs that may be involved, and on issues that must be considered, when organizing and managing these two types of virtual communities effectively.
Our objective in doing so is not only to provide guidance to marketers in managing their internet venues, but also to stimulate academic researchers to consider these issues in depth.

2. Theoretical background and hypothesis

In developing a theory of consumer participation in virtual communities, one approach has been to postulate that a number of individual-level and group-level variables act separately to influence the consumer’s desires, we-intentions, and ultimately his or her participation in the community (B&D; see also Bagozzi, 2000). An alternative perspective, which builds upon this view, and one that we adopt in this article, is that, whereas both individual-level and group-level variables are important drivers of virtual community participation, at least some of the individual-level variables are antecedents to group-level variables, which in turn influence participation. Such a perspective is consistent with social identity theory (Hogg & Abrams, 1988) as well as recent research on online social interactions (McKenna & Bargh, 1999) and views group influences on the participant to stem from an explicit understanding that group membership yields beneficial outcomes. Using this approach, we start with a set of individual-level motives that help explain why consumers participate in virtual communities. To the extent that these motives can be satisfied through participation, the community should exert influence on its members. Our theoretical model (see Fig. 1) is developed in detail next.

2.1. Individual motives for participation in the virtual community

To understand the motives of virtual community participants, we draw upon the well-established uses and gratifications paradigm, originally developed and employed by communications researchers to understand people’s motivations for using different media (e.g., Flanagan & Metzger, 2001; McQuail, 1987). This research has shown that individuals often seek

![Fig. 1. A social influence model of virtual community participation.](image-url)
out media in a goal-directed fashion to fulfill a core set of motivations, which are also helpful in understanding why consumers might participate in virtual communities.

Of special relevance from a marketing perspective, informational value is one that the participant derives from getting and sharing information in the virtual community, and from knowing what (presumably credible) others think. We also included instrumental value that a participant derives from accomplishing specific tasks, such as solving a problem, generating an idea, influencing others regarding a pet issue or product, validating a decision already reached or buying a product, through online social interactions (e.g., Hars & Ou, 2002; McKenna & Bargh, 1999). These objectives are all instrumental in the sense that they are usually defined prior to participation and facilitate achievement of specific end-state goals (Bagozzi & Dholakia, 1999).

Although informational and instrumental values tend to be viewed as distinct by communication researchers (e.g., Flanagin & Metzger, 2001), it is perhaps more appropriate to view them as constituents of a single purposive value construct from a marketing perspective, which we define as “the value derived from accomplishing some pre-determined instrumental purpose” (including giving or receiving information) through virtual community participation. Indeed, the empirical analyses reported below support this reformulation.

The second type of value, self-discovery, involves understanding and deepening salient aspects of one’s self through social interactions. One aspect of self-discovery is to interact with others so as to obtain access to social resources and facilitate the attainment of one’s future goals (McKenna & Bargh, 1999). Another aspect of self-discovery is that such interactions may help one to form, clearly define and elaborate on one’s own preferences, tastes, and values. Whereas purposive value relates to utilitarian concerns connecting one’s self to external objects or issues, self-discovery focuses on intrinsic concerns, constituted by or embedded in the self itself. But both these values are self-referent, i.e., they primarily involve and refer to one’s personal self.

The next two values we included have more to do with others, i.e., other members of the virtual community. Maintaining interpersonal connectivity refers to the social benefits derived from establishing and maintaining contact with other people such as social support, friendship, and intimacy. Several studies have shown that many participants join such communities mainly to dispel their loneliness, meet like-minded others, and receive companionship and social support (e.g., McKenna & Bargh, 1999; Wellman & Gulia, 1999). Social enhancement is the value that a participant derives from gaining acceptance and approval of other members, and the enhancement of one’s social status within the community on account of one’s contributions to it (Baumeister, 1998). Studies have shown that many participants join virtual communities mainly to answer others’ questions and to provide information, for recognition by peers (Hars & Ou, 2002).

Maintaining interpersonal connectivity and social enhancement both emphasize the social benefits of participation, and are group-referent, i.e., the referent of these values is the self in relation to other group members. This distinction between self- and group-referent values is important, since later on, we develop the idea that the type of virtual community dictates which values are more influential in predicting social influence and participation therein.

Finally, the last value we included is entertainment value, derived from fun and relaxation through playing or otherwise interacting with others. Studies have shown that many participants do so for entertainment through exploring different fictional identities, encountering, and solving virtual challenges, etc. (McKenna & Bargh, 1999).

2.2. Social influences on member participation in the virtual community

In their model, B&D (2002) hypothesized that three group-level influences drive virtual community participation: compliance (i.e., normative influence of others’ expectations), internalization (i.e., congruence of one’s goals with those of group members), and identification (i.e., conception of one’s self in terms of the group’s defining features). B&D found that internalization and identification were significant predictors of participation, but compliance was not. This non-significant result for compliance is not surprising since participation in virtual communities is usually voluntary and anonymous, and members are
able to leave without much effort. So most members may not feel the need to comply with others’ expectations. We did not include compliance influences in our model, instead viewing identification and internalization to be the two salient social influences of the virtual community on member participation.

Such a two-factor view of social influence is favored by existing sociological research as well (e.g., McMillan & Chavis, 1986; Postmes et al., 2000; Wellman, 1999). For instance, Etzioni (1996) suggests that two characteristics are necessary for a social grouping to be considered a community. First, a community requires an understanding of, and a commitment by the individual to, a sense of values, beliefs, and conventions shared with other community members, i.e., internalization. Second, it entails a web of affect- and value-laden relations (of varying strengths) among a group of individuals, often reinforcing one another, and going beyond the immediate utilitarian purpose of a particular interaction, i.e., identification with the group.

2.2.1. Social identity in the virtual community

Social identity captures the main aspects of the individual’s identification with the group in the sense that the person comes to view himself or herself as a member of the community, as “belonging” to it. This is a psychological state, distinct from being a unique and separate individual, conferring a shared or collective representation of who one is (Hogg & Abrams, 1988), and involves cognitive, affective, and evaluative components (e.g., Bergami & Bagozzi, 2000; Ellemers et al., 1999). In a cognitive sense, social identity is evident in categorization processes, whereby the individual forms a self-awareness of virtual community membership, including components of both similarities with other members, and dissimilarities with non-members (Ashforth & Mael, 1989; Turner, 1985).

Belonging to a virtual community also has emotional and evaluative significance (Tajfel, 1978). In an emotional sense, social identity implies a sense of emotional involvement with the group, which researchers have characterized as attachment or affective commitment (e.g., Bagozzi & Dholakia, 2002; Ellemers et al., 1999). Emotional social identity fosters loyalty and citizenship behaviors in group settings (e.g., Bergami & Bagozzi, 2000; Meyer, Stanley, Herscovitch & Topolnytsky, 2002), and is useful in explaining consumers’ willingness to maintain committed relationships with firms in marketing settings (Bhattacharya & Sen, 2003). Finally, since the definition of one’s identity influences one’s sense of self-worth (e.g., Blanton & Christie, 2003), social identity also entails an evaluative component. Evaluative social identity is measured as the individual’s group-based or collective self-esteem and is defined as the evaluation of self-worth on the basis of belonging to the community. In our model, the cognitive, affective, and evaluative elements are components of a second-order social identity construct (see Fig. 1).

Identifying with a virtual community that one has chosen volitionally stems from an understanding that membership entails significant benefits. Consistent with this view, social identity theorists posit that identification with social groups is derived, first and foremost, from their functionality—groups are identified with to the extent that they fulfill important needs of the member (Hogg & Abrams, 1988). While some needs may concern the self alone, others may also be group-referenced. Based on this discussion, we hypothesize that:

**Hypothesis 1.** Higher levels of value perceptions lead to a stronger social identity regarding the virtual community.

2.2.2. Group norms in the virtual community

Internalization, operationalized here by group norms, refers to the adoption of common self-guides for meeting idealized goals shared with others, because they are viewed as coinciding with one’s own goals. It may therefore be defined as an understanding of, and a commitment by, the individual member to a set of goals, values, beliefs, and conventions shared with other group members. Group norms are especially relevant for virtual communities since they are perhaps the most readily accessible (for instance, through FAQs) or inferable (from archives of previous interactions, for example) elements of group-related information available in many communities (Postmes et al., 2000) and regulating interactions among members over time (Alon et al., 2004).

Group norms become known to members in different ways. One occurs upon joining the community, where the new participant actively seeks out the
group’s goals, values, and conventions. In other cases, the participant slowly comes to discover the community’s norms through socialization and repeated participation therein, over a period of time. A third possibility is that the individual learns of the community’s norms beforehand and joins the community on account of one’s perceived overlap with the community’s norms. In order to be influential, group norms should be volitionally accepted by members as congruent to their own motives (Postmes et al., 2000). An understanding of what one seeks to gain from participation should be a crucial antecedent to group norms. Therefore,

**Hypothesis 2.** Higher levels of value perceptions lead to stronger group norms regarding the virtual community.

In addition to providing knowledge regarding what the community’s objectives are and how it interacts together, an understanding and acceptance of its group norms by itself allows the individual to consider oneself as its full-fledged member. Because of this, once the member has learnt and accepted the virtual community’s norms, he or she will identify with the community more. In this regard, Hogg and Abrams (1988) note that cooperative interdependence resulting from the pursuit of shared goals results in the establishment of a well-defined group structure—which in turn leads its members to identify with it. Similarly, in a virtual community context, Alon et al. (2004) postulate that instrumental behaviors and the understanding of each others’ goals precede the establishment and propagation of the community’s identity, in their model of community life cycles. Hence,

**Hypothesis 3.** Stronger group norms lead to a stronger social identity regarding the virtual community.

Hypothesis 3 implies that value perceptions influence social identity in two ways: directly and also through their impact on group norms. Next, it is useful to consider the specific processes by which group norms advance the individual’s desires for participation.

At one level, strong group norms implicitly generate consensus among members regarding when and how to engage in online social interactions. In this respect, group norms promote mutual agreement among group members regarding the specific details of participation itself. In a second sense, research on group negotiation has shown that group norms facilitate a cooperative motivational orientation among group members (Weingart, Bennett, & Brett, 1993). Philosopher Bratman (1997) similarly notes that shared intentional activity is preceded by associated forms of mutual responsiveness on the members’ parts to do whatever it takes to be able to complete their own parts in enabling joint action to occur. Group norms should therefore increase participants’ inclinations to mutually accommodate their schedules, preferences and commitments with others’ in order to be able to engage in group action. Thus,

**Hypothesis 4.** Stronger group norms lead to stronger mutual agreement to participate in the virtual community.

**Hypothesis 5.** Stronger group norms lead to a stronger willingness to mutually accommodate each other to enable participation.

Both mutual agreement and mutual accommodation represent mechanisms through which the participant moves from rather general and broadly defined goals and conventions of the group, toward actualizing specific episodes of online social interactions. In this sense, they serve as mediators by which group norms influence the individual’s participation desires in our model. Both provide the potential for deciding to engage in virtual community activities but do not, in and of themselves, necessarily provide the motivation to do so. The transformation of mutual agreement and accommodation into intentions to engage in virtual community activities is hypothesized to be provided by felt desires to engage in these activities. Desires provide the motivation to decide in favor of acting as part of a virtual community. Therefore,

**Hypothesis 6.** Stronger mutual agreement leads to stronger desires to participate in the virtual community.

**Hypothesis 7.** Stronger mutual accommodation leads to stronger desires to participate in the virtual community.

At the same time, we posit that participation desires are also influenced by social identity. Since identification renders a person to maintain a positive self-defining relationship with other virtual community members, he or she will be motivated to engage in
behaviors needed to do so (Hogg & Abrams, 1988). An important part of maintaining this relationship with the group is to actively participate in online social interactions. In this respect, social identities prescribe and instigate group-oriented behaviors. As examples, Ellemers et al. (1999) studied experimentally formed groups and found that aspects of social identity influenced acts of in-group favoritism, whereas Bergami and Bagozzi (2000) found that social identity led to performance of organizational citizenship behaviors by firm employees. Based on this discussion,

**Hypothesis 8.** Stronger social identity leads to stronger desires to participate in the virtual community.

Consistent with the B&D model, we view desires as mediators of the influence of individual and group-level antecedents on we-intentions. Since we study intentional social action, the referent of the participant’s actions is the virtual community rather than one’s self. A we-intention is defined as a “commitment of an individual to engage in joint action and involves an implicit or explicit agreement between the participants to engage in that joint action” (Tuomela, 1995, p. 9; see B&D, 2002 for a detailed discussion). We note here that such joint action may not necessarily be contemporaneous; members can perform their respective parts at different times. Nevertheless, joint actions entail coordinated endeavors between group members.

The role played by desires is to transform the multiple reasons for acting found in the antecedents, which in our model are individual and social reasons for participating, into an overall motivation to act. Since such behavior is effortful, involving a greater or lesser degree of effort (e.g., remembering when to meet or respond to a group member, adjusting other engagements in one’s schedule to interact online, etc.), desires are necessary precursors to we-intentions in performing such actions (Perugini & Bagozzi, 2001). Based on this discussion,

**Hypothesis 9.** Stronger desires lead to higher levels of we-intentions to participate in the virtual community.

Further, we posit that the mediation of desires in the effects of the social influence variables on we-intentions will be partial. This is because, participation in virtual communities, although goal-directed, involves both effortful as well as habitual components. The habitual aspects of such actions are relevant since many members may have belonged to the virtual community for a long time beforehand, having developed routines of participation therein. For many participation episodes, behavior may be automatic, as in checking periodically to see if new messages have been posted on a bulletin-board to which one belongs. For such habitual participation, group norms and social identity should influence we-intentions directly, rather than through desires, depending on the strength of one’s habit. Therefore,

**Hypothesis 10.** Stronger group norms lead to higher levels of we-intentions to participate in the virtual community.

**Hypothesis 11.** A stronger social identity leads to higher levels of we-intentions to participate in the virtual community.

Finally, it is important to stress that whereas the B&D (2002) analysis ended with we-intentions, we also measured participants’ behaviors in a second wave, expecting we-intentions to significantly predict subsequent participation, in accordance with standard attitude-theoretic formulations (Eagly & Chaiken, 1993). Hence,

**Hypothesis 12.** Higher levels of we-intentions lead to higher levels of participation in the virtual community.

2.3. Network-based and small-group-based virtual communities

In the literature on virtual communities (and especially so within marketing), they have tended to be construed as vast, vaguely defined, social spaces comprised of ever-changing congregations of participants (e.g., B&D, 2002; Wellman et al., 1996; Williams & Cothrel, 2000). The implicit assumption in such construals is that this abstract social category, the community as a whole, is the salient basis of social identity and group norms for all its members. Such a view also does not allow one to distinguish between different types of communities that might meet in different internet venues (see Section 3.1 below), nor does it allow for the possibility that the nature of the community may change over time as repeated
interactions among members result in the formation of interpersonal relationships (Alon et al., 2004). While internet venues such as bulletin-boards and chat rooms may be unambiguous to organizers or outside observers, their participants may have starkly different views regarding who belongs to the virtual communities located therein, what their values are, and how central they are for its members.

In exploring this issue further, the distinction made by sociologists (e.g., Wellman, 1999) between neighborhood solidarities, defined as tightly bounded, densely knit groups with strong relationships between members, and social networks, defined as loosely bounded, sparsely knit networks of members sharing narrowly defined relationships with one another, is useful. Whereas neighborhood solidarities are geographically conjoint groups, where each member knows everyone else and relies on them for a wide variety of social support, social networks are usually geographically dispersed groups that interact with one another for a specific reason, and usually without prior planning (Wellman, 1999).

Social psychologists similarly distinguish between common bond and common identity groups (Prentice, Miller, & Lightdale, 1994; Sassenberg, 2002). Whereas bonds between members are the glue holding the group together in common bond groups, such attachment is dependent on identification to the whole group, in common identity groups. Common bond groups therefore correspond to neighborhood solidarities, whereas common identity groups correspond to social networks. This distinction, of viewing the community as either the same group of individuals with each of whom the person has relationships, or viewing it as a venue where people (strangers or acquaintances) with shared interests or goals meet, provides a useful typology of marketing relevance to classify virtual communities.

In some instances, the member’s definition of the virtual community may primarily be in terms of the venue, and only superficially associated with any particular individuals within it. For instance, a person may log into a bulletin-board on gardening, and participate because he is interested in the subject-matter, but have no expectation or inclination to meet, chat or communicate with any particular individual therein. Similarly, an engaged Amazon.com customer may read and benefit from reviews offered by other customers, without any personal knowledge of, or relationships with, the reviewers. We call a virtual community defined this way, i.e., a specialized, geographically dispersed community based on a structured, relatively sparse, and dynamic network of relationships among participants sharing a common focus, to be a network-based virtual community.

In other cases, the member may identify primarily with a specific group (or groups) of individuals, rather than with the online venue itself. For example, a software developer may log on to a messaging system specifically to chat with her geographically distant “buddy group” of software developers every Wednesday night to trade ideas, learn new concepts, and socialize. Here, the developer’s focus is on communication with peers that she knows personally, rather than on the venue of the AOL messaging system. We call such a virtual community, constituted by individuals with a dense web of relationships, interacting together online as a group, in order to accomplish a wider range of jointly conceived and held goals, and to maintain existing relationships, to be a small-group-based virtual community. These are virtual communities because they meet through online venues for a significant proportion (but not necessarily all) of their overall interactions together, as a group. Moreover, they often have commercial focuses. For example, within such company-sponsored organizations as Harley Owners Groups (HOGs), many small-group-based virtual communities exist that participate extensively in internet-based activities, which are augmented by face-to-face interactions periodically.

2.4. The moderating role of community type in the social influence model

We first consider how members’ motivations for participation might vary between these two virtual communities. To do so, it is useful to better understand how small-group-based and network-based communities differ from each other. One important difference between them is that the specific group with which the member interacts holds greater importance for those belonging to small-group-based when compared to network-based communities. This is because the individual knows everyone else personally, and may
often have special shared histories and close personal relationships with them. As a result, relationships between group members are likely to be stronger, more resilient, and more stable than those in network-based communities, where members are more likely to participate primarily to achieve functional goals (e.g., to learn how to install a software program) and may have tenuous, short-lived, and easily severed relationships with others.

Accentuating the importance of the group for small-group-based virtual community members is also the fact that the particular virtual community is often only one of a number of places where such groups meet. Online social interactions are often supplemented by face-to-face and other offline forms of interactions. For instance, a small group of HOG members may not only chat online with one another periodically in the course of a week, but meet on weekdays for coffee and fellowship, and on weekends for group outings. In contrast, network-based virtual community members are more likely to interact with each other exclusively online.

These differences all point to the greater importance of group-referent values, for small-group-based community members and self-referent values for network-based virtual community members. As a result, we expect that:

**Hypothesis 13.** Purposive and self-discovery value perceptions will be stronger for network-based when compared to small-group-based virtual community members.

**Hypothesis 14.** Maintaining interpersonal connectivity and social enhancement will be stronger for small-group-based when compared to network-based virtual community members.

These posited distinctions in strength of value perceptions should manifest themselves in differences in the expressed mean levels of value perceptions by members of the two virtual communities. Further, we expect that, since these different motivations—self-referent for network-based and group-referent for small-group-based virtual communities—provide the impetus for participation, they should also influence the social influence variables, group norms, and social identity to a much greater extent, respectively. Specifically, we expect that,

**Hypothesis 15.** The impact of purposive and self-discovery values on group norms and social identity will be stronger for network-based than for small-group-based virtual community members.

**Hypothesis 16.** The impact of maintaining interpersonal connectivity and social enhancement on group norms and social identity will be stronger for small-group-based than for network-based virtual community members.

Taken together, all the above hypotheses provide an understanding of why consumers participate in virtual communities, the bases of the community’s social influence, as well as differences between small-group-based and network-based virtual communities.

### 3. Empirical study

#### 3.1. Finding members of small-group-based and network-based virtual communities

As noted, online venues offer a useful starting point for finding both types of virtual community members. For the sake of generalizability, we included virtual communities from seven different types of internet venues (e.g., Catterall & Maclaran, 2001) in this study. The first type, email lists, refers to specialized mailing lists organized around particular topics of interest, and are widely used by firms to maintain customer relationships. An example in our study included the DisneyDollarLess Club for budget-minded tourists. The second type, website bulletin boards, is company-sponsored venues, where participants can post and read messages about the firm’s products and services. An example in our study included the “Advanced Squad Leader” website. The third type was Usenet newsgroups, each having a specific focus of interest such as technical issues (e.g., Linux installation), hobbies (e.g., Pokemon), and specific products and brands (e.g., Ford

---

1 This website was sponsored and maintained by Multi-Man Publishing, publisher of the Advanced Squad Leader video game.
Mustang cars). Among others, our study included members from the alt.marketing.ebay and alt.guitar-amps newsgroups.

The fourth venue was real-time online-chat systems, such as ICQ and AOL instant messenger, both of which were represented in our study. These venues allow participants to chat with others in real time. The fifth type of venue was web-based chat rooms such as those on the AOL and MSN websites. Examples in our study included the AOL Word Haven chat room and the NHB chat room on pork.com. The sixth type of venue we included was multiplayer virtual games, wherein gamers can play as a group by simultaneously logging online together, through wired or wireless interfaces. Examples of networked video games in our study included Diablo II, Dungeon Siege, and Neverwinter Nights. Finally, the seventh venue included in the study was multi-user domains (MUDs). MUDs are a special form of real-time computerized conferencing, where participants don pseudonymous personas and role play in quests, masquerades, games, and also in work-related communal interactions (Wellman et al., 1996). Among examples of MUDs included in our study were Avatar, Wheel of Time, and Xyllomer.

3.2. Pre-test

To better understand whether these seven venues harbor small-group-based communities, network-based communities or both, we conducted a pretest with 240 regular participants in these venues. Participants were first asked to choose the venue that they participated in most often, and then to describe their interactions therein in detail. These descriptions were content-analyzed by two coders. Specifically, each response was coded into one of the following three categories: (1) the respondent usually interacts with the same group of people; (2) the respondent usually interacts with different individuals or groups of people; and (c) unable to determine the type of interaction. Of the 240 decisions made, the 2 coders agreed on 213 (or 89%) decisions. The remaining decisions were resolved after comparison and discussion. After eliminating those descriptions which could not be gauged by the coders for interaction type, the final classification can be found in Table 1, which provides the proportion of respondents by type of venue indicating that they participated either with the same group or with different groups every time.

The results showed that most participants of the first three venues—email lists, website bulletin boards, and Usenet newsgroups—engaged in interactions with different individuals or groups on each occasion. In contrast, a vast majority of participants in the remaining four venues interacted with the same group on most occasions. Based on these results, we concluded that the first three venues would be the most suitable for finding network-based virtual communities, whereas the last four venues would be appropriate for finding small-group-based virtual communities for our study.

3.3. Method of main study

We then collected data from regular participants in the seven venues by conducting an internet-based

<table>
<thead>
<tr>
<th>Virtual community venue</th>
<th>Proportion participating with same group every time</th>
<th>Proportion participating with different groups every time</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email listsa</td>
<td>3.4%</td>
<td>96.6%</td>
<td>29</td>
</tr>
<tr>
<td>Website bulletin boardsa</td>
<td>17.6%</td>
<td>82.4%</td>
<td>34</td>
</tr>
<tr>
<td>Usenet newsgroups</td>
<td>3.0%</td>
<td>97.0%</td>
<td>33</td>
</tr>
<tr>
<td>Real-time online-chat systemsb</td>
<td>97.2%</td>
<td>2.8%</td>
<td>34</td>
</tr>
<tr>
<td>Web-based chat roomsb</td>
<td>90.0%</td>
<td>10.0%</td>
<td>20</td>
</tr>
<tr>
<td>Multiplayer virtual gamesb</td>
<td>100%</td>
<td>0%</td>
<td>6</td>
</tr>
<tr>
<td>Multi-user domainsb</td>
<td>85.2%</td>
<td>14.8%</td>
<td>27</td>
</tr>
</tbody>
</table>

a Determined as suitable for finding networks.
b Determined as suitable for finding small groups.
survey in the Spring of 2002. The survey was publicized by contacting a significant number of organizers of popular online venues of each type. These organizers informed their membership about the survey, and encouraged their members to participate by visiting a website where the survey was made available.

The study was introduced to participants as an “opinion survey regarding group interactions on the internet.” Participants were asked to select the venue that they most frequently visited when online, giving them the opportunity to complete the survey regarding the type of virtual community with which they were most familiar. After this selection was made, participants described their chosen interaction in some detail such as the name of the venue, the date when they first joined, whom they normally interacted with, details regarding their interactions, what they liked about their online group, etc.

Based on our pretest results, participants of the four venues corresponding to small-group-based virtual communities were then branched to another section of the survey, where they were told: “Imagine that you are logging on to the internet to engage in the group interaction that you described above. You have a number of friends within that group that you regularly interact with. Please picture briefly in your mind the name and image of each online friend. Then write your first name and their first names/handles in the table below. You may include up to, but not necessarily, five group members. Please be sure to include only friends that are part of the group you regularly interact with on the internet.”

Similarly, since our pretest results indicated that most participants of the remaining three venues interacted with whomever was online, they were then branched to a section where they described their last online interaction in detail. These respondents were then told to visualize up to five average members, using them as representatives of the other virtual community members. All participants, regardless of the venue selected, responded to the same set of measures.

3.4. Sample characteristics and measures

A total of 545 participants representing 264 different virtual communities completed the survey. Of the entire sample, 41.8% were female, 54.3% were male, while 3.9% did not disclose their gender. Respondents ranged in age from 18 to 79, with a mean age of 33.1 years (median=30, S.D.=13.43). While 387 (71%) were US residents, the other 29% belonged to a total of 27 other countries. Canada (n=42, 7.7%), Australia (n=23, 4.2%), and Germany (n=21, 3.9%) were the three next largest subgroups, represented in the sample. On average, respondents had been online for 7.53 years (S.D.=3.57), suggesting a high level of experience.

All of the measures used in the survey are provided in Table 2. The value perception measures were the same as those used by Flanagin and Metzger (2001), and were introduced with the following preface, “How often do you use your online group (as identified above) for satisfying the following needs?” The measures of group norms, social identity, desires, and we-intentions were similar to those used by B&D (2002).

Because of the large number of different virtual communities (264) involved, we measured participation behaviors through self-reports, rather than other means such as observation. Participants were contacted through a follow-up email approximately a month later to obtain this information, with two reminders to encourage responses given thereafter. A total of 465 (or 85.3%) participants responded to this second-wave of questions regarding participation behavior with response rates ranging from 80.9% to 91.4% depending on venue type.

3.5. Preliminary analysis

Our full sample model includes participants of all seven venues and is used to test Hypotheses 2–12, and our network-based and small-group-based subsamples are used to test the moderation (Hypotheses 13–16). All of the models (CFA and SEM) described below were run using the LISREL 8.52 program (Jöreskog & Sörböm, 1999). The goodness-of-fit of the models was assessed with chi-square tests, the root mean square error of approximation (RMSEA), the non-normed fit index (NNFI), and the comparative fit index (CFI). Discussions of these indices can be found in Bentler (1990), Browne and Cudeck (1993), Marsh and Hovecar (1985), and Marsh, Balla, and Hau (1996). Satisfactory model fits are indicated by non-
Table 2
Details of measures in the main study

Constructs and measures

<table>
<thead>
<tr>
<th>Constructs and measures</th>
<th>( \rho_e )</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purposive value</strong> (nine measures)</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To get information</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To learn how to do things</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To provide others with information</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To contribute to a pool of information</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To generate ideas</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To negotiate or bargain</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To get someone to do something for me</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To solve problems</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
<tr>
<td>To make decisions</td>
<td>( \rho_e = 0.84^{\text{b}} )</td>
<td>0.75</td>
</tr>
</tbody>
</table>

| **Self-discovery value** (two measures)                                                | \( \rho_e = 0.89 \) | 0.69 |
| To learn about myself and others                                                      | \( \rho_e = 0.89 \) | 0.69 |
| To gain insight into myself                                                           | \( \rho_e = 0.89 \) | 0.69 |

| **Maintaining interpersonal interconnectivity** (two measures)                        | \( \rho_e = 0.94 \) | 0.76 |
| To have something to do with others                                                   | \( \rho_e = 0.94 \) | 0.76 |
| To stay in touch                                                                      | \( \rho_e = 0.94 \) | 0.76 |

| **Social enhancement value** (two measures)                                           | \( \rho_e = 0.89 \) | 0.66 |
| To impress                                                                            | \( \rho_e = 0.89 \) | 0.66 |
| To feel important                                                                     | \( \rho_e = 0.89 \) | 0.66 |

| **Entertainment value** (four measures)                                               | \( \rho_e = 0.90 \) | 0.71 |
| To be entertained                                                                     | \( \rho_e = 0.90 \) | 0.71 |
| To play                                                                               | \( \rho_e = 0.90 \) | 0.71 |
| To relax                                                                              | \( \rho_e = 0.90 \) | 0.71 |
| To pass the time away when bored                                                      | \( \rho_e = 0.90 \) | 0.71 |

| **Cognitive social identity** (two measures)                                          | \( \rho_e = 0.87 \) | 0.68 |
| Please indicate to what degree your self-image overlaps with the identity of the group of friends as you perceive it (seven-point “not at all–very much” scale) | \( \rho_e = 0.87 \) | 0.68 |
| How would you express the degree of overlap between your personal identity and the identity of the group you mentioned above when you are actually part of the group and engaging in group activities? (eight-point “not at all–very much” scale) | \( \rho_e = 0.87 \) | 0.68 |
| **Affective social identity** (two measures)                                          | \( \rho_e = 0.87 \) | 0.68 |
| How attached are you to the group you mentioned above? (seven-point “not at all–very much” scale) | \( \rho_e = 0.87 \) | 0.68 |
| How strong would you say your feelings of belongingness are toward the group you mentioned above? (seven-point “not at all–very much” scale) | \( \rho_e = 0.87 \) | 0.68 |

Table 2 (continued)

<table>
<thead>
<tr>
<th>Constructs and measures</th>
<th>( \rho_e )</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaluative social identity</strong> (two measures)</td>
<td>( \rho_e = 0.97 )</td>
<td>0.84</td>
</tr>
<tr>
<td>“I am a valuable member of the group” (seven-point “agree–disagree” scale)</td>
<td>( \rho_e = 0.97 )</td>
<td>0.84</td>
</tr>
<tr>
<td>“I am an important member of the group” (seven-point “agree–disagree” scale)</td>
<td>( \rho_e = 0.97 )</td>
<td>0.84</td>
</tr>
</tbody>
</table>

| **Group norms** (two measures)                                                         | \( \rho_e = 0.92 \) | 0.71 |
| “Interacting together sometime within the next 2 weeks with your online group can be considered to be a goal. For each of the people listed below, please estimate the strength to which each holds the goal” (five-point “weak–strong” scales) | \( \rho_e = 0.92 \) | 0.71 |

| **Strength of self’s goal**                                                            | \( \rho_e = 0.93 \) | 0.72 |
| Average of the strength of group members’ goal                                         | \( \rho_e = 0.93 \) | 0.72 |

| **Mutual agreement** (two measures)                                                    | \( \rho_e = 0.97 \) | 0.80 |
| “How strong would you say the explicit or implicit agreement is among each of the following to interact with on the internet as a group sometime during the next 2 weeks? (five-point “weak–strong” scales) | \( \rho_e = 0.97 \) | 0.80 |

| **Strength of self’s agreement**                                                       | \( \rho_e = 0.93 \) | 0.80 |
| Average of the strength of group members’ agreement                                    | \( \rho_e = 0.93 \) | 0.80 |

| **Mutual accommodation** (two measures)                                                | \( \rho_e = 0.94 \) | 0.93 |
| “How willing are each of the following to accommodate or adjust to the needs of the others in the group so as to choose a time and place to interact together on the internet sometime during the next 2 weeks? (five-point “not at all willing–very willing” scales) | \( \rho_e = 0.94 \) | 0.93 |

| **Strength of self’s willingness to accommodate**                                      | \( \rho_e = 0.94 \) | 0.93 |
| Average of the strengths of group members’ willingness to accommodate                  | \( \rho_e = 0.94 \) | 0.93 |

| **Desires** (three measures)                                                           | \( \rho_e = 0.94 \) | 0.93 |
| “I desire to interact with the group sometime during the next 2 weeks” (seven-point “agree–disagree” scale) | \( \rho_e = 0.94 \) | 0.93 |
| “My desire for interacting together with the group can be described as” (seven-point “very weak desire–very strong desire” scale) | \( \rho_e = 0.94 \) | 0.93 |
| “I want to interact together with my group during the next 2 weeks.” (seven-point “does not describe me at all–describes me very much” scale) | \( \rho_e = 0.94 \) | 0.93 |
significant chi-square tests, RMSEA $\leq 0.08$, and NNFI and CFI values $\geq 0.90$.

Two indicators were used to operationalize each latent construct in the CFA and the SEM. For latent constructs where more than two items were available (informational value, instrumental value, entertainment value, and desires), these were combined to produce two indicators according to the so-called “partial disaggregation model” (Bagozzi & Edwards, 1998). Compared to models where every item is a separate indicator, this yielded models with fewer parameters to estimate, and reasonable ratios of cases to parameters, while smoothing out measurement error to a certain extent. All analyses were performed on covariance matrices (Cudeck, 1989). An initial exploratory analysis and examination of the correlation matrix showed that the correlations between the measures of informational value and instrumental value were very high. Consequently, and because such a combination is theoretically justifiable (see our earlier discussion), these two values were treated as a single construct labeled “purposive value” with four measures, two each of informational and instrumental value.

3.6. Results

3.6.1. Measurement model evaluation

We evaluated the internal consistency and discriminant validity of model constructs. Given space considerations, the results for only the full sample are reported here in detail. The results for the subsamples were substantively similar and are available from the authors.

3.6.2. Internal consistency

We used two measures to evaluate internal consistency of constructs. The composite reliability ($\rho_c$) is a measure analogous to coefficient $\alpha$ (Bagozzi & Yi, 1988; Fornell & Larcker, 1981, Eq. (10)), whereas the average variance extracted ($\rho_{VE}$) estimates the amount of variance captured by a construct’s measure relative to random measurement error (Fornell & Larcker, 1981, Eq. (11)). Estimates of $\rho_c$ above 0.60 and $\rho_{VE}$ above 0.50 are considered supportive of internal consistency (Bagozzi & Yi, 1988). The $\rho_c$ and $\rho_{VE}$ values for all constructs in the model (provided in Table 2) were significantly higher than the stipulated criteria, and therefore indicative of good internal consistency.

3.6.3. Discriminant validity

Discriminant validity of the model constructs was evaluated using three different approaches. A confirmatory factor analysis model was built with 14 latent constructs and a total of 29 measures. Results showed that the model fit the data well. The goodness-of-fit statistics for the model were as follows: $\chi^2(287)=1010.83$, $p \approx 0.00$, RMSEA=0.07, SRMR=0.04, NNFI=0.95, CFI=0.96. The $\phi$-matrix (correlations between constructs, corrected for attenuation) is provided in Table 3. As a first test of discriminant validity, we checked whether the correlations among the latent constructs were significantly less than one. Since none of the confidence intervals of the $\phi$-values (± two standard errors) included the value of one (Bagozzi & Yi, 1988), this test provides evidence of discriminant validity.

Secondly, for each pair of factors, we compared the $\chi^2$-value for a measurement model constraining their correlation to equal one to a baseline measurement model without this constraint. A $\chi^2$-difference test was performed for each pair of factors (a total of 91
tests in all), and in every case resulted in a significant difference, again suggesting that all of the measures of constructs in the measurement model achieve discriminant validity.

Third, we performed a test of discriminant validity suggested by Fornell and Larcker (1981). This test is supportive of discriminant validity if the average variance extracted by the underlying construct is larger than the shared variance (i.e., the \( \phi^2 \) value) with other latent constructs. This condition was satisfied for all of the 91 cases. In sum, internal consistency and discriminant validity results enabled us to proceed to estimation of the structural model.

3.7. Structural model estimation

Structural models were built separately for the full sample (to test Hypotheses 1–12), as well as for the network and small group subsamples (to test Hypotheses 13–16). Table 4 provides the goodness-of-fit statistics for these models and \( R^2 \) values of the endogenous constructs. Tests of mediation and comparisons with rival models were conducted on the full sample to test its robustness. Using multiple-sample analyses in LISREL, structured means analyses were conducted to test Hypotheses 13 and 14, and tests of moderation were conducted to test Hypotheses 15 and 16.

3.7.1. Full sample model

Considering the fit-statistics from Table 4, the chi-square is significant \( (p<0.05) \), which is usually the case for large sample sizes. All the other statistics are within the acceptable ranges for the full model, indicating a good fit to the data. Considering social identity first, both purposive \( (\gamma=0.15, \text{S.E.}=0.07) \) and entertainment \( (\gamma=0.21, \text{S.E.}=0.07) \) values are significant predictors of social identity, whereas the other three value perceptions are not, supporting Hypothesis 1. Examining the antecedents of group norms next, two of the five value perceptions, purposive \( (\gamma=0.34, \text{S.E.}=0.10) \) and self-discovery \( (\gamma=0.19, \text{S.E.}=0.09) \) values, have significant paths to group norms, whereas the other three do not. Twenty-four percent of the variance in group norms is explained by value perceptions, supporting Hypothesis 2. Fig. 2 summarizes these and subsequent results.

In examining Hypotheses 3–5 which explicate the associations between group norms and its consequences, we find that group norms influences social identity \( (\beta=0.23, \text{S.E.}=0.07) \), mutual agreement \( (\beta=0.83, \text{S.E.}=0.07) \), and mutual accommodation \( (\beta=0.77, \text{S.E.}=0.07) \), providing support to all three
hypotheses. Sixty-two percent, 57% and 43% of variance in social identity, mutual agreement, and mutual accommodation are explained by their antecedents, respectively.

Considering whether these variables influence desires to participate next, we find that mutual agreement ($\beta=0.34$, S.E.=0.07) and social identity ($\beta=0.59$, S.E.=0.13) do influence desires, but mutual accommodation does not ($\beta=-0.05$, S.E.=0.05). Thus, Hypotheses 6 and 8 are supported, but Hypothesis 7 is not. Thirty-seven percent of the variance in desires is explained by these antecedents. On hindsight, the non-significant effect of mutual accommodation on desires is perhaps not surprising, since for many of the participants belonging to network-based communities and interacting with different members.

Table 4
Goodness-of-fit statistics for structural models

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Full sample</th>
<th>Small-group-based virtual communities subsample</th>
<th>Network-based virtual communities subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>$\chi^2(342)=927.77, p&lt;0.001$</td>
<td>$\chi^2(342)=854.25, p&lt;0.001$</td>
<td>$\chi^2(342)=824.73, p&lt;0.001$</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.072</td>
<td>0.074</td>
<td>0.074</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.061</td>
<td>0.067</td>
<td>0.066</td>
</tr>
<tr>
<td>NNFI</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>CFI</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

$R^2$ values for endogenous variables

<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>Small-group-based virtual communities subsample</th>
<th>Network-based virtual communities subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group norms</td>
<td>0.24</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>Social identity</td>
<td>0.62</td>
<td>0.85</td>
<td>0.47</td>
</tr>
<tr>
<td>Mutual accommodation</td>
<td>0.43</td>
<td>0.35</td>
<td>0.32</td>
</tr>
<tr>
<td>Mutual agreement</td>
<td>0.57</td>
<td>0.42</td>
<td>0.56</td>
</tr>
<tr>
<td>Desires</td>
<td>0.37</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td>We-intentions</td>
<td>0.54</td>
<td>0.58</td>
<td>0.66</td>
</tr>
<tr>
<td>Participation behavior</td>
<td>0.24</td>
<td>0.33</td>
<td>0.21</td>
</tr>
</tbody>
</table>

We-intentions:

<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>Small-group-based virtual communities subsample</th>
<th>Network-based virtual communities subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Social Identity</td>
<td>1.00</td>
<td>86***</td>
<td>82***</td>
</tr>
<tr>
<td>Affective Social Identity</td>
<td>1.59</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Evaluative Social Identity</td>
<td>21***</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Social Identity</td>
<td>$R^2=.42$</td>
<td>$R^2=.43$</td>
<td>$R^2=.43$</td>
</tr>
<tr>
<td>Mutual Agreement</td>
<td></td>
<td>$R^2=.57$</td>
<td></td>
</tr>
<tr>
<td>Mutual Accommodation</td>
<td></td>
<td>$R^2=.43$</td>
<td></td>
</tr>
<tr>
<td>Desires</td>
<td></td>
<td>$R^2=.37$</td>
<td></td>
</tr>
<tr>
<td>We-intentions</td>
<td></td>
<td>$R^2=.54$</td>
<td></td>
</tr>
<tr>
<td>Participation Behavior</td>
<td></td>
<td>$R^2=.24$</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Parameter estimates for final structural model. Unstandardized coefficients and standard errors in parentheses; insignificant paths are omitted for ease of exposition.
every time, given the mutual agreement to participate, the necessity for mutual accommodation to adjust to the needs of others may not be an issue. Instead, they may be willing to interact with whomever is online.

Considering the direct impact of group-influence variables on we-intentions, the path from group norms (β=0.43, S.E.=0.06) is significant, but that from social identity (β=0.16, S.E.=0.10) is not. Thus, Hypothesis 9 is supported but Hypothesis 10 is not. Finally, supporting Hypotheses 11 and 12, the paths from desires to we-intentions (β=0.19, S.E.=0.04), and from we-intentions to behavior (measured in the second-wave; β=1.43, S.E.=0.19) are both significant. Fifty-four percent of the variance in we-intentions and 17% of the variance in behavior is explained by their antecedents.

3.7.2. Tests of mediation

To obtain further support for the validity of the model, rather than using a saturated model where “everything is related to everything” as the baseline, we performed formal tests of mediation for all possible paths in our model. This was done to check if additional direct paths not included in the model were significant. Specifically, we conducted 7 tests to check for the significance of a total of 32 potential paths. As an example, to check if the direct paths from the five value perceptions to desires were significant, we compared the model described above with a model in which five additional direct paths were added from the five value perceptions to desires. The difference in chi-square values between the two models (χ²(5)=6.37), with five degrees of freedom, is a test of the significance of these added paths. Since this difference is not significant (p>0.27) and none of the individual paths is significant, we concluded that the direct paths from the value perceptions to desires are insignificant, and therefore group norms and social identity mediate all of the effects of value perceptions on desires, as hypothesized.

Of 32 potential paths tested, results show that only 3 of these were significant. The direct paths from entertainment value to behavior, entertainment value to mutual agreement, and mutual agreement to we-intentions were all significantly greater than zero (dashed arrows in Fig. 2). The goodness-of-fit statistics for a model including these three paths were as follows: χ²(339)=884.21, p=0.00, CFI=0.96, NNFI=0.95, RMSEA=0.069, and SRMR=0.058. The R² values of the three endogenous variables after incorporating these additional significant paths were mutual agreement (0.57 vs. 0.57 before), we-intentions (0.54 vs. 0.54 before), and behavior (0.24 vs. 0.17 before). The other 29 paths were not significant, providing additional evidence that our proposed model is robust, and suggesting that the social influence variables mediate most of the effects of value perceptions on participation in virtual communities.

3.7.3. Moderating effects of virtual community type (test of Hypotheses 13–16)

We conducted multiple sample analyses (Jöreskog & Sörbom, 1999) for the network and small-group subsamples to test the hypotheses regarding the moderating role of virtual community type. Table 5 provides the means, standard deviations, and the Cronbach α reliabilities of the constructs for the subsamples. As is evident, the reliabilities are good overall.

Hypothesis 13 posited that the self-referent values, i.e., purposive and self-discovery values, would be greater for the network-based relative to the small-group-based subsample, whereas Hypothesis 14 posited that the group-referent values, maintaining interpersonal connectivity and social enhancement, would be greater for the small-group-based relative to the network-based virtual communities. To test these hypotheses, we conducted a structured means analysis in LISREL, using the following model of means structures (Jöreskog & Sörbom, 1999): \( x(g)=\tau_x+\Lambda_x\xi(g)+\delta(g) \), where \( g=\text{small-group} \) and \( x(g) \) is a vector of input variables, \( \tau_x \) is a vector of constant intercept terms, \( \Lambda_x \) is a matrix of coefficients of the regression of \( x \) on \( \xi \), \( \xi \) is a vector of constant independent variables, \( \delta \) is a vector of measurement errors in \( x \) and the means of the \( \xi(g)\)\(=\kappa(g)\).

We set the \( \kappa(\text{small-group})=0 \) to define the origin and units of measurement of the \( \xi \) factors and computed \( \kappa(\text{network}) \), and then determined whether the differences in the factor means of the two groups were significantly different from each other. Table 6 provides the results.

\[ \text{Footnote: The detailed results are available from the authors upon request.} \]
As can be seen, the factor mean of purposive value was significantly higher for the network subsample, but that of self-discovery was not different between the two groups. These results partially support Hypothesis 13. Considering the two group-referent values, from Table 6, we find that both, maintaining interpersonal connectivity and social enhancement factor means were higher for the small-group relative to the network subsample, providing support to Hypothesis 14. Interestingly, entertainment value although not characterized as either self- or group-referent, was also significantly higher for the small group subsample, suggesting that this value perception might have a group-referent basis. These results provide evidence that the motivations of participants in the two virtual communities have different bases of reference.

To test Hypotheses 15 and 16, we conducted tests of moderation to determine whether the strengths of the paths from value perceptions to social identity and group norms were different between the small-group-based and the network-based subsamples. Table 7 summarizes the analyses and results.

Consider the first test presented in Table 7. To test Hypothesis 15, that the purposive value to group norms path is stronger for network-based relative to small group-based virtual communities, we ran two multiple-sample models. In the first model, all paths were unconstrained between the two groups. This is the no constraints model in the first row of Table 7. In the second model, the purposive value to group norms path was constrained to be equal for both subsamples. This is the equal paths model.

The difference in chi-square values between the two

Table 5
Means, standard deviations, and reliabilities for construct measures

<table>
<thead>
<tr>
<th>Scale</th>
<th>Small-group-based virtual community subsample (N=278)</th>
<th>Network-based virtual community subsample (N=265)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of</td>
<td>Mean</td>
</tr>
<tr>
<td>Purposive value</td>
<td>9</td>
<td>25.43</td>
</tr>
<tr>
<td>Self-discovery</td>
<td>2</td>
<td>5.49</td>
</tr>
<tr>
<td>Maintain inter. interconnectivity</td>
<td>2</td>
<td>7.01</td>
</tr>
<tr>
<td>Social enhancement</td>
<td>2</td>
<td>4.03</td>
</tr>
<tr>
<td>Entertainment value</td>
<td>4</td>
<td>15.88</td>
</tr>
<tr>
<td>Group norms</td>
<td>2</td>
<td>7.07</td>
</tr>
<tr>
<td>Cognitive SI</td>
<td>2</td>
<td>6.97</td>
</tr>
<tr>
<td>Affective SI</td>
<td>2</td>
<td>10.16</td>
</tr>
<tr>
<td>Evaluative SI</td>
<td>2</td>
<td>10.25</td>
</tr>
<tr>
<td>Mutual agreement</td>
<td>2</td>
<td>7.56</td>
</tr>
<tr>
<td>Mutual accommodation</td>
<td>2</td>
<td>6.78</td>
</tr>
<tr>
<td>Desires</td>
<td>3</td>
<td>15.89</td>
</tr>
<tr>
<td>We-intentions</td>
<td>2</td>
<td>7.71</td>
</tr>
<tr>
<td>Participation behavior</td>
<td>1</td>
<td>6.38</td>
</tr>
</tbody>
</table>

Table 6
Test of factor mean differences between network and small group subsamples for the five value perception constructs (Hypotheses 13 and 14)

<table>
<thead>
<tr>
<th>Value perceptions</th>
<th>Small-group-based sub-sample factor mean$^a$ $\mu_{(small-group)}$</th>
<th>Network-based sub-sample factor mean $\mu_{(network)}$</th>
<th>t-value (significance level)</th>
<th>Result of statistical est.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purposive value</td>
<td>0</td>
<td>0.29</td>
<td>4.21 ($p&lt;0.001$)</td>
<td>$\mu_{(small-group)}&lt;\mu_{(network)}$</td>
</tr>
<tr>
<td>Self-discovery value</td>
<td>0</td>
<td>-0.16</td>
<td>-1.76 (ns)</td>
<td>$\mu_{(small-group)}=\mu_{(network)}$</td>
</tr>
<tr>
<td>Maintaining interpersonal connectivity</td>
<td>0</td>
<td>-0.61</td>
<td>-6.23 ($p&lt;0.001$)</td>
<td>$\mu_{(small-group)}&lt;\mu_{(network)}$</td>
</tr>
<tr>
<td>Social enhancement value</td>
<td>0</td>
<td>-0.31</td>
<td>-4.03 ($p&lt;0.001$)</td>
<td>$\mu_{(small-group)}&lt;\mu_{(network)}$</td>
</tr>
<tr>
<td>Entertainment value</td>
<td>0</td>
<td>-1.01</td>
<td>-12.55 ($p&lt;0.001$)</td>
<td>$\mu_{(small-group)}&lt;\mu_{(network)}$</td>
</tr>
</tbody>
</table>

$^a$ Reference value: factor means set to zero.
models (χ²(1)=2.80) with a single degree of freedom, provides a test of the equality of the path for the two groups. Since this difference approaches significance at the p ≈ 0.09 level, we may conclude that the direct path between purposive value and group norms is marginally greater for network- when compared to small-group-based virtual communities. Other tests were conducted similarly.

As can be seen from Table 7, in testing Hypothesis 15, three of the paths are significantly greater for the network- relative to the small-group-subsample, and one path (self-discovery value to social identity) is not. These findings are supportive of Hypothesis 15.

The results for Hypothesis 16 are mixed. The paths from maintaining interpersonal connectivity to group norms and social identity are significantly higher for small-group-based versus network-based virtual communities. However, the paths from social enhancement to neither of these variables are significantly different for the two subsamples. Hypothesis 16 thus...

### Table 7

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Path coefficients in unconstrained model</th>
<th>Chi-square statistics test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purposive value → group norms is greater for network than for small group.</td>
<td>γ(network)=0.32a (0.14b), γ(small-group)=0.01 (0.12)</td>
<td>Equal paths model: χ²(497)=1678.31; test of H₁: χ²(1)=2.80, p ≈ 0.09; difference marginally significant</td>
</tr>
<tr>
<td>Self-discovery value → group norms is greater for network than for small group.</td>
<td>γ(network)=0.38 (0.12), γ(small-group)=0.04 (0.11)</td>
<td>Equal paths model: χ²(497)=1679.72; test of H₁: χ²(1)=4.21, p&lt;0.05; difference is significant</td>
</tr>
<tr>
<td>Purposive value → social identity is greater for network than for small group.</td>
<td>γ(network)=0.37 (0.15), γ(small-group)=−0.03 (0.15)</td>
<td>Equal paths model: χ²(497)=1679.29; test of H₁: χ²(1)=3.78, p=0.05; difference is significant</td>
</tr>
<tr>
<td>Self-discovery value → social identity is greater for network than for small group.</td>
<td>γ(network)=0.33 (0.13), γ(small-group)=−0.17 (0.13)</td>
<td>Equal paths model: χ²(497)=1676.31; test of H₁: χ²(1)=0.80, p=0.37; difference not significant</td>
</tr>
<tr>
<td>Omnibus test: for all four paths together</td>
<td></td>
<td>Equal paths model: χ²(500)=1692.51; test of H₁: χ²(4)=17.00, p&lt;0.001; all four paths combined are different between the subsamples</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Path coefficients in unconstrained model</th>
<th>Chi-square statistics test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypothesis 15</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintaining interpersonal connectivity → group norms is greater for small group than for network</td>
<td>γ(network)=−0.11 (0.11), γ(small-group)=0.15 (0.10)</td>
<td>Equal paths model: χ²(497)=1678.30; test of H₁: χ²(1)=2.79, p ≈ 0.09; difference marginally significant</td>
</tr>
<tr>
<td>Social enhancement value → group norms is greater for small group than for network.</td>
<td>γ(network)=−0.02 (0.13), γ(small-group)=0.19 (0.10)</td>
<td>Equal paths model: χ²(497)=1676.53; test of H₁: χ²(1)=1.02, p=0.31 Difference not significant</td>
</tr>
<tr>
<td>Maintaining interpersonal connectivity → social identity is greater for small group than for network</td>
<td>γ(network)=−0.06 (0.12), γ(small-group)=0.46 (0.13)</td>
<td>Equal paths model: χ²(497)=1684.41; test of H₁: χ²(1)=8.90, p &lt;0.05; difference is significant</td>
</tr>
<tr>
<td>Social enhancement value → social identity is greater for small group than for network.</td>
<td>γ(network)=0.02 (0.14), γ(small-group)=−0.30 (0.12)</td>
<td>Equal paths model: χ²(497)=1677.42; test of H₁: χ²(1)=1.91, p&lt;0.16; difference not significant</td>
</tr>
<tr>
<td>Omnibus test: for all four paths together</td>
<td></td>
<td>Equal paths model: χ²(500)=1687.35; test of H₁: χ²(4)=11.84, p&lt;0.01; all four paths combined are different between the subsamples</td>
</tr>
</tbody>
</table>

a Unstandardized coefficient.  
b Standard error.
receives support for maintaining interpersonal connectivity but not for social enhancement value.

4. General discussion

Our empirical survey-based study, conducted across a variety of different virtual communities, found overall support for our proposed social influence model of virtual community participation. The findings suggest that an appropriate conceptualization of intentional social action in virtual communities is one where the community’s influence on members stems from an understanding or expectation of various benefits that participants seek to attain from social interactions therein. Further, we also found that there are interesting differences between network-based and small-group-based virtual community participants in both levels of self-referent and group-referent motives, as well as their impacts on the social influence variables. These findings raise several interesting issues discussed below.

4.1. Understanding how to deliver value desired by virtual community participants

For participants of network-based virtual communities, purposive value was found to be a key driver of participation. From a managerial perspective, such purposive motives can be characterized as complementary to each other. For instance, in measuring informational value, one item that we used was “to get information,” whereas another one was “to provide information to others.” It can be argued that an information-seeker will find the virtual community useful only if he or she can find another participant with the complementary motive of providing that information. As a result, an important task of network-based virtual community managers may be defined in terms of matching of participants’ complementary motives effectively and maintaining a balance so that the purposive goals of most participants are achieved.

The finding that social benefits such as maintaining interpersonal connectivity and social enhancement are significant drivers of participation in small-group-based virtual communities is also noteworthy. Since it suggests that many participants in such communities are interested in engaging in social interactions together, as a group, the marketer’s objective may be defined in terms of matching group members’ preferences to interact together.

These differences also imply that virtual community organizers will need to thoughtfully decide on which tools and functionalities to provide in their venues. In network-based virtual communities, members may find “applications of purpose”—tools, application, and content that enables them to achieve their goals successfully—to be valuable. Examples of such applications include comprehensive FAQs lists, organization of past responses from community members in transparent and easily accessible hierarchies, query-tools to match information-seekers to information providers, and so on.

In contrast, in small-group-based communities, “applications of process” that enable uninterrupted, vivid and enjoyable group interaction may be more valuable. Aesthetic and easily learnt user interfaces, the ability to engage in interactive communication (see Section 4.2), and tools allowing members to contact and solicit the participation of group members, are all examples of applications of process. Understanding the relative importance and effectiveness of different applications of process and applications of purpose that enable self- and group-referent motives to be attained through future research seems especially important in future research.

4.2. Venue characteristics and virtual community type

In our pretest, we found internet venues to strongly favor either network- or small-group-based virtual communities, suggesting that managers might be able to influence the type of communities that they organize by offering specific venues to their consumers. A closer examination of these venues suggests that the type of communication processes therein—whether interactive or non-interactive—play an important role in determining whether network- or small-group-based virtual communities dominate.

The degree of interactivity of communication processes within a virtual community is influenced by both, the synchronicity of communication—the capability of enabling a participant to formulate and deliver a response in real time, allowing a real-time dialogue to occur (Hoffman & Novak, 1996), and by
the number and range of inputs that the participant can provide, such as text, audio, video, etc. (Lombard, 2001). Whereas communication researchers have studied the impacts of synchronicity and input attributes on the outcomes of communication processes (e.g., Lombard, 2001), we know relatively little about the marketing-relevant impacts of these variables for virtual communities.

There are several reasons why higher levels of interactivity may be more suitable for small-group-based and lower levels of interactivity more conducive to network-based virtual communities. First, researchers have shown that a high level of interactivity generally entails a higher level of involvement on the participant’s part (Hoffman & Novak, 1996). This implies that in interactions involving high interactivity, stronger relationships between participants may be necessary; in addition, participants should be responsive and engaged throughout the duration of the interaction—such requisites all the more characteristic of small groups.

Second, the greater the interactivity afforded by the venue, the higher the likelihood of spontaneity between the participants, the more the possibility of interruption or preemption, and the greater the mutuality and patterns of turn-taking (Brown & Yule, 1983; Zack, 1993). Again, interactions of this type are possible when participants know and understand each other well.

Third, it is more likely that interactive exchanges will continue and be repeated at future times when participants can engage in many different topics of conversation, move easily from one topic to the next, and have at least some shared history or knowledge base, all requiring broad-based relationships.Confirming this prediction, as well as studying the specific influences of interactivity and its constituents on the economic activities within virtual communities, are promising future research issues.

4.3. Understanding how to convey member information to other participants

The mode by which the identity and information about a member is conveyed to others is also likely to be influenced by community type. In network-based communities, because members don’t know each other at first in most cases, and their motives are self-referent, a member’s reputation is likely to be crucial as a means of establishing trust and status and for fostering social interactions (Resnick, Zeckhauser, Friedman & Kuwabara, 2000; Rheingold, 2002). Reputation mechanisms considering contribution frequency and quality made may therefore need to be carefully and elaborately designed for such communities. Communities such as slashdot also choose to leave a visible trail of each member’s contribution history for other members to see and judge them.

On the other hand, because small-group-based community members know each other well and participate to achieve group-referent goals, reputation systems may not be required or may be less essential. Instead, in this case, it may prove more useful to enable members to share a detailed personal, self-composed history with other members—for instance, through “About Me” web-pages or in-depth member profiles. On the whole, we know relatively little about the importance of different information elements of a participant’s reputation or other identifying information, and when or how such measures are used by participants to make interaction decisions within the virtual community (see Rheingold, 2002 for a detailed discussion). These questions offer interesting future research opportunities.

4.4. Concluding thoughts

Through studying the antecedents of social influence, and making and elaborating on the distinction between network- and small group-based communities, our broad objective in this exploratory study was to stimulate thinking among practitioners and researchers regarding the scope of virtual communities for marketing applications. Our contention is that marketers for the most part have tended to view virtual communities narrowly, focusing entirely on network-based communities. Through our presentation, we defined and elaborated on a second type, the small-group-based virtual community, found empirical differences, and developed some practically useful distinctions between them. The following two issues deserve more elaboration.

First, more development is needed into expanding the conceptual difference between venues and virtual communities. For example, it is important to note that
the same venue, such as a slashdot bulletin-board, may possibly host both network- and small-group-based virtual communities at the same time. Indeed, even a particular person may belong to both virtual communities within this venue as when she exchanges ideas frequently with her regular group of friends weekly, yet sometimes reads messages on bulletin-boards to update her knowledge on current software issues. But as we noted, high levels of interactivity and other features such as applications of process, imply that certain venues are more conducive for small-group-based virtual communities, and others for network-based virtual communities. Indeed, marketing managers may be able to influence the type of communities that are built within their venue through an informed selection of the venue characteristics discussed above.

Second, as noted before, it is possible, indeed very likely that some groups within network-based virtual communities may over time evolve into small-group-based virtual communities, as frequent interactions among the same individuals result in greater knowledge and the building of interpersonal relationships (see Alon et al., 2004 for a detailed discussion). But until we learn more about the conditions leading to such transitions, it will be difficult to draw conclusions or make inferences regarding a particular online group’s future type, based on its current type. Building on our consumer-centric definition presented earlier, the monitoring and management of a virtual community is best viewed as an ongoing task by its organizers.

A final point we wish to make is to summarize the differences between our model and the B&D (2002) model. Decision making in our model is a direct function of social influence and an indirect function (through social influence) of value perceptions, whereas decision making in the B&D model is a direct function of both social influence and individual-level variables. Our model therefore makes stronger predictions in the sense that a particular sequence is specified among social- and individual-level antecedents, whereas these are left as exogenous predictors in the B&D model. Second, our model proposes five specific categories of value perceptions derived from the communications literature, whereas B&D rely upon general, summary variables derived from the theory of planned behavior (i.e., attitudes, subjective norms, perceived behavioral control) and the model of goal-directed behavior (i.e., positive and negative anticipated emotions). Our antecedents have more managerial relevance than those found in B&D. Third, unlike B&D, we developed explanations based on contingencies inherent in different types of virtual communities. Finally, our tests of the model went a step beyond B&D’s tests by including participation behavior as a dependent variable.

In spite of these contributions, it is important to recognize the exploratory nature of this research, and its attendant limitations. For instance, of the five hypothesized benefits, two—maintaining interpersonal interconnectivity and social enhancement—did not have significant effects on any of the variables. This suggests that more research is needed to determine all of the benefits, and differences between the two communities in this regard. In conclusion, it seems important to echo the optimism expressed by marketing scholars studying virtual communities (e.g., Balasubramanian & Mahajan, 2001), and suggest that these online forums are only likely to grow in importance, influence, and the activities for which they are used, as consumers become more comfortable and acclimatized with these environments and marketers learn how to forecast, monitor, and design their communication programs to take advantage of such opportunities. They merit continued and increasing attention from both practitioners and academicians.

Acknowledgements

We would like to thank the participants of the “Defining the Value of Virtual Communities” special session at the 2002 AMA Summer Educators Conference for their comments. Special thanks to the editor and three anonymous reviewers.

References


Corporate social responsibility and consumers’ attributions and brand evaluations in a product–harm crisis

Jill Klein¹,², Niraj Dawar³,⋆,²

¹INSEAD, Fontainebleau, France
²Richard Ivey School of Business, University of Western Ontario, London, Ontario, Canada N6A 3K7

Abstract

A growing body of research on Corporate Social Responsibility (CSR) in marketing has shown that (1) CSR plays a role in consumers’ brand and product evaluations, over and above economic or ‘rational’ considerations such as product attributes; and (2) CSR has a spillover or ‘halo effect’ on otherwise unrelated consumer judgments, such as the evaluation of new products. Yet CSR’s halo on consumer behavior may extend beyond product evaluations, into nonroutine types of judgments such as attributions. We examine the possibility that the CSR halo affects consumers’ attributions in a product–harm crisis situation. In two studies that employ experimental manipulations of prior CSR on a sample of consumers, we examine whether attributions that are influenced by CSR mediate the impact of product–harm crises on consumers’ brand evaluations. The results of Study 1 support the hypothesis. Study 2 introduces a boundary condition on the results of Study 1, showing that mediation effects are only found for consumers that are CSR-sensitive. The findings point to a role of CSR in consumer behavior that is more complex than previously conceptualized.

Keywords: Corporate social responsibility; Product–harm crises; Attributions; Brand evaluations

Corporate Social Responsibility (CSR) has emerged in recent years as both an important
Mclaughlin, 1996; Russo & Fouts, 1997; Waddock & Smith, 2000). From a marketing perspective, the firm’s economic benefits from CSR have been documented in its link to consumers’ positive product and brand evaluations, brand choice, and brand recommendations (Brown & Dacin, 1997; Drumwright, 1994; Handelman & Arnold, 1999; Osterhus, 1997; Sen & Bhattacharya, 2001). Through a variety of theoretical lenses, the important contributions of these studies have been to demonstrate that (1) CSR plays a role in routine consumer behavior, over and above economic or ‘rational’ considerations such as product attributes; and (2) CSR has a spillover or ‘halo effect’ on otherwise unrelated routine consumer judgments, such as the evaluation of new products.

Yet the CSR halo may also play a very different role in consumer behavior. The CSR halo may influence consumer judgments in nonroutine settings that are rarely studied. In these settings, CSR may operate for the firm as an insurance policy against the negative impact of untoward events such as product–harm crises. In these nonroutine settings, the most common consumer cognitive process is attribution (Folkes, 1984). Yet the impact of CSR on attributions remains unresearched. Attributions are important because they form the basis of revision and updating of enduring and central consumer judgments such as brand evaluations (Folkes, 1988). The objective of the present research is to uncover this potentially broader scope of CSR in consumer behavior, and to map its impact on consumer attributions of blame in a product–harm crisis setting, and through them on to brand evaluations and purchase intentions. The general hypothesis tested is that CSR affects the attribution process itself, and that attributions in turn influence brand evaluations. Thus, CSR associations are cast as a moderator of attribution. These attributions are conceptualized as a mediator of the impact of product–harm crisis on brand evaluations. The product–harm crisis setting is a particularly opportune one for the study of brand evaluations, because not only does it activate corporate associations such as CSR, but the sudden and often large changes in brand evaluation that accompany product–harm crises also make such a setting an invaluable opportunity for researchers to understand influences on brand evaluations.

1. Conceptual background

1.1. The corporate social responsibility halo

A halo effect is the ‘bias’ due to a measure that spills over to another measure (Thurstone, 1920). For example, a strong consumer belief about the performance characteristics of a Porsche may spill over onto beliefs about its reliability; or a consumers’ overall attitude toward a brand might spillover onto their assessment of specific attributes of that brand (Beckwith & Lehmann, 1975). In this study, we are specifically interested in the halo effect of consumers’ prior beliefs about the company’s position on CSR, onto attributions about a product–harm crisis involving that company. The halo effect is most conspicuous when the affected measures are unrelated to the source of the halo, because when the affected measures are related, it is not possible to separate the halo effect from a nomological effect. Previous studies in marketing have shown how CSR affects unrelated product evaluations. In this study, we examine how CSR beliefs spillover onto attributions made about a product–harm crisis.

Brown and Dacin’s (1997) definition of CSR as the organization’s status and activities with respect to its perceived societal obligations, provides a useful starting point that has also been adopted in subsequent research on CSR in marketing (e.g., Sen & Bhattacharya, 2001). As in these previous studies, we adhere to this definition but focus on consumers’ perceptions of CSR as drivers of consumer behavior. This general definition of CSR allows for many different operationalizations, but the corporate record on the environment is one of the most frequently used manifestations of the construct (Bansal & Roth, 2000; Klassen & Mclaughlin, 1996; Osterhus, 1997), and is one of the six key dimensions on which SOCRATES, the influential CSR index published by KLD Research, rates companies. Accordingly, our lens on CSR is that of consumers’ perceptions of corporate environmental responsibility.

CSR associations are an element of overall ‘corporate associations,’ distinct from attribute-level information about products (Aaker, 1996). Product associations, or stored information about product attributes, are generally used as input for consumers to make everyday product-related judgments such as
evaluations of quality, comparisons with other products, and assessments of value for money. In the interest of cognitive economy, consumers are likely to use information beyond product associations only sparingly, and mainly when existing information is insufficiently diagnostic to make the judgment at hand (Feldman & Lynch, 1988). Even so, CSR associations have been shown to have a spillover effect on product and brand evaluations. In a nonroutine setting such as a product–harm crisis, corporate associations including CSR are all the more likely to be activated, and therefore, we expect the CSR halo to exert an influence on judgments such as attributions (Brown & Dacin, 1997).

1.2. Product–harm crises and attributions

Product–harm crises have been defined as well-publicized instances of defective or dangerous products (Dawar & Pillutla, 2000). The increasing complexity of products, more demanding customers, and more vigilant media are making product–harm crises an ever more visible occurrence. Recent crises involving brands such as Coca Cola in Europe, Firestone in the United States, and Snow milk in Japan, have created consumer and media awareness and sensitivity to such crises. Product–harm crises can imperil long-standing favorable customer impressions about the brand. Firms often institute expensive voluntary product recalls to minimize damage to their brands from product–harm crises. Even so, much of the loss of market capitalization that is associated with product recalls is due to the damage to intangibles such as consumers’ perceptions of the brand, rather than by the costs of the recall itself (Davidson & Worrell, 1992; Pruitt & Peterson, 1986). The seriousness and frequency of product–harm crises contrasts with the relative paucity of research in this area.

We know that consumers spontaneously construct attributions of blame for faulty or harmful products (Folkes, 1984; Folkes & Kotsos, 1986). These attributions are important from a marketing perspective because they form the basis of consumers’ brand judgments and behavior. In constructing these attributions consumers rely on information, including corporate associations, that goes beyond the product attributes that are normally the basis of evaluation or purchase decisions (Aaker, 1996; Folkes, Koletsky, & Graham, 1987). The impact of these associations on attributions may depend on their relevance to consumers (Crocker, 1980; Metalsky & Abramson, 1981); high importance placed on CSR issues may lead consumers to utilize such information in forming their attributions. Weiner’s (1980) widely used attribution model conceptualizes three causal dimensions of attribution that lead to an overall judgment of responsibility or blame: (1) the locus of the behavior (the event that triggers the crisis), which can be internal or external to the actor (in our case, the firm); (2) the stability of the behavior, which can be unchanging or temporary; and (3) the controllability of the behavior, which can be within or outside the control of the actor. If the locus is internal, and the behavior is stable and controllable, observers (in our case, consumers) tend to attribute responsibility to the actor, and subsequent consumer behavior such as blame or anger, is directed toward the actor. If on the other hand, the locus is external, and the behavior is temporary and uncontrollable, attributions will tend to be made to external factors (Folkes, 1984). The recent product–harm crisis involving Firestone tire blowouts that allegedly caused consumer deaths, and the subsequent recall of millions of tires, helps illustrate this model of the attribution process. If consumers believe that the tires were poorly made, that Firestone has had a history of product defects, and that they could have averted the problems with better quality control, they will be likely to attribute responsibility to Firestone. In contrast, if they believe the problems were caused by harsh driving or vehicle conditions, that this is the first time Firestone’s tires have been implicated as the cause of accidents, and that driving and vehicle conditions are in fact outside the control of the tire manufacturer, they will be more likely to attribute responsibility to external factors, such as to the vehicle manufacturer or to driving conditions. Our intention in studying attributions in a product–harm crisis setting is to establish whether CSR influences locus, stability, and controllability attributions.

1.3. Attributions and consumer behavior

Kelley and Michela (1980) in their broad review of attribution theory characterized the field as consisting of studies of the antecedents or determinants of
attributions, such as information, beliefs and motivation, and the consequences, such as affect, behavior, and expectancy. Marketing studies of attribution have tended to focus on consequences rather than informational antecedents. Folkes (1984), in a pioneering study, demonstrated that consumer reactions to product failure are a function of the values on the three causal dimensions of locus, stability, and controllability. Variance in consumers’ desire for a refund or exchange, their expectation of an apology, and even their desire for revenge were predicted on the basis of the values on the three causal dimensions. In a series of subsequent studies, Folkes et al. (1987) and Folkes and Kotsos (1986) showed how attributions in the context of service delays led to the subsequent desire to complain, and affected repurchase behavior. Jorgensen (1994) applied Weiner’s model to consumers’ attributions in the context of serious company disasters (a fatal airliner crash), and found that consumers’ attributions of the cause of the incident changed their affect and attitudes toward the company. Taken together, these studies provide compelling evidence for the effects of attributions on consumers’ attitudes and behavior. But less is known about the determinants of attributions. Studies of informational antecedents have found mixed results (e.g., Sparkman & Locander, 1980; Yalch & Yoshida, 1983). Sparkman and Locander (1980) reported that the context changed consumers’ attributions about celebrity product endorsement. However, despite manipulating a number of contextual cues, they found that only the extent to which other celebrities also endorsed the advertised product was a significant predictor of attributions. Yalch and Yoshida (1983) also found little evidence of contextual antecedents on attributions. One explanation suggested for the mixed results is that these studies employed realistic stimuli to increase external validity, but that these stimuli introduced strong prior beliefs that swamped the effects of experimental manipulations (Folkes, 1988). Additionally, the antecedents examined in these studies are in fact contextual cues that are closely related to the behavior or actor, and the studies did not examine locus, stability, or controllability as contributing factors in the attribution. In contrast, our interest lies in understanding the effects of information such as CSR associations on attributions of stability, locus, and controllability. Furthermore, unlike these previous studies, we seek to systematically vary prior CSR beliefs to determine their impact on attributions on all three causal dimensions. Finally, to ensure that we accounted for any dampening effects of realistic stimuli, we ran two studies, one using a fictitious firm as the actor, and a second a real firm.

In Weiner’s (1980) model, the three causal dimensions of attribution are not objective facts based on unbiased data, but rather judgments formulated on the basis of information about the event or behavior available to the observer. In the context of product–harm crises, such information may be gleaned from the media, from the company, and from other sources. Attributions are potentially prone to influence because they are constructed through an interaction of event-related information with the observer’s prior beliefs (Folkes, 1988). Indeed, biases due to prior beliefs have been shown to influence consumer judgments in product harm crises. Dawar and Pillutla’s (2000) data showed that consumers’ interpretation of a firm’s response to a product–harm crisis was subject to their prior expectations about the firm. Expectations were defined in terms of consumers’ accumulated experience with the company and information about its past behavior toward customers. Their research focuses on the interaction of firm response with consumers’ prior expectations. While it provides a basis for positing directional effects, it does not touch on the effects of specific corporate associations such as CSR, nor does it address the potentially important mediating role of attributions.

Our hypotheses are drawn from an integration of these strands of research on the impact of CSR on consumer behavior, on consumers’ attributions, and the impact of product–harm crises on brand evaluations. We propose that CSR beliefs will be activated in response to a product–harm crisis, as part of the activation of corporate associations that occurs because consumers engage in making attributions about the crisis. This activation enhances the likelihood of the CSR halo having a spillover effect on attributional judgments. Specifically, we hypothesize that CSR beliefs moderate consumers’ perception of the locus of the crisis event as internal or external, whether they see it as stable or temporary, and whether they believe it to have been controllable or not. CSR beliefs are a key element of the knowledge
of the company that an individual holds and draws upon to make these kinds of judgments (Brown & Dacin, 1997). As with other types of halo effects, we expect that information related to the crisis will be interpreted in a confirmatory fashion (e.g., Darley & Gross, 1983; Snyder & Swann, 1978). In particular, we propose that for firms that enjoy positive prior CSR, the trigger event for the product–harm crisis will be judged as more external, less stable and less controllable, relative to firms that do not enjoy positive CSR. Specifically, we hypothesize that:

**H1a.** The locus of the crisis will be perceived as external rather than internal when prior CSR is positive vs. when prior CSR is negative;

**H1b.** The crisis event will be perceived as unstable rather than stable when prior CSR is positive vs. when prior CSR is negative; and

**H1c.** The crisis event will be perceived as uncontrollable rather than controllable when prior CSR is positive vs. when prior CSR is negative.

In addition, as in previous studies of attribution, we anticipate that blame for the event will be influenced by attributions (Folkes & Kotsos, 1986), but because we view CSR as a moderator of attributions, we suggest that attributions will mediate the relationship between CSR and blame for the firm.

**H2.** Attributions will mediate the relationship between CSR and company blame.

Blame, in turn, will affect consumers’ brand evaluations, over and above the direct effects of CSR on brand evaluation. That is, positive or negative CSR information is likely to directly affect brand evaluations, but blame is expected to also affect brand evaluations, even controlling for the direct effect of CSR.

Following the logical consequences, we know brand evaluations affect consumers’ purchase intentions. In other words, we expect blame to be inversely correlated with brand evaluations, and brand evaluations to be positively related to purchase intentions. This hypothesis is intended as a validation of the knock-on effects of the attributions. The existence of these effects will indicate that the effects on attribution do indeed lead to potentially enduring changes in brand evaluations, and that these in turn have an effect on behavioral intentions. Specifically:

**H3a.** Blame will be inversely related to brand evaluations.

**H3b.** Brand evaluations will predict buying intentions.

2. Study 1

2.1. Design and measures

The design involved three between-subjects conditions (positive CSR, negative CSR, and a control condition in which no information about CSR was provided), and accordingly, three versions of the questionnaire. In the present study, a fictitious firm name was used. At the outset respondents were provided with background information about the firm that included the following description:

The following information is about a real, well-known oil company that has been in business for many decades. For the purposes of this study we will call the company OilCo.

In the positive CSR condition, subjects read that OilCo had been ranked 1st among 14 major oil companies on treatment of the environment, that the company was viewed as very environmentally responsible, had placed at the top of similar environmental rankings in the past, and had consistently shown that it cared about the environment. Furthermore, OilCo’s recent work on the preservation of the Peruvian rain forest was cited. In the negative condition, respondents read that OilCo had been ranked last of 14 major oil companies on treatment of the environment, that the company was viewed as very environmentally irresponsible, had been placed at the bottom of similar environmental rankings in the past, and had consistently shown that it did not care about the environment. Furthermore, OilCo’s ‘recent harm caused to the Peruvian rain forest’ was cited. In the negative condition, respondents read that OilCo had been ranked last of 14 major oil companies on treatment of the environment, that the company was viewed as very environmentally irresponsible, had been placed at the bottom of similar environmental rankings in the past, and had consistently shown that it did not care about the environment. Furthermore, OilCo’s ‘recent harm caused to the Peruvian rain forest’ was cited. In the control condition, no information was given about OilCo’s environmental record.

After reading the introduction, respondents (in the treatment conditions) rated OilCo on environmental social responsibility, as a manipulation check. They
also indicated how important it is that a company is environmentally responsible. Respondents then read a newspaper article in which an OilCo product failure was described:

**OilCo Lubricant Tied to Engine Problems**

**[Dateline]** There have been several hundred reports of severe engine damage linked to OilCo’s ET Synthetic Blend Lubricant, an automotive engine lubricant designed to protect engines and improve performance.

Apparently, ingredients in the oil interact with the plastic jugs in which the lubricant is sold. Over time, the chemicals in the plastic thicken the oil, which causes engine damage. The company that makes the plastic jugs is Carson Plastics.

An OilCo spokesperson said yesterday, “There is no problem with the lubricant if it is sold by stores before the expiration date marked on the bottle. All of the engine problems encountered can be tied to a national grocery store chain that sold the lubricant after the expiration date”.

After reading the article, respondents completed the questionnaire that included questions concerning the locus, stability and controllability of the problem, company blame, brand evaluations, buying intention, and several demographic questions. Questions were asked in this order (with the exception that the brand evaluation and buying questions were mixed together). The specific items were consistent with previous research on attributions and brands (Agarwal & Rao, 1996; Dawar & Pillutla, 2000; Weiner, 1980).

2.2. **Respondents and procedure**

One hundred and fifty respondents participated in a mall–intercept study. The mall was located in the suburbs of a large city in a Midwestern state and attracted predominantly middle-class shoppers. The average age of the sample was 36.7 (ranging from 17 to 74) and 64% were female. Ninety-eight percent had at least a high school diploma, with 44% of the sample holding a college degree. Questionnaires were self-administered and respondents were recruited and randomly assigned to condition by professional interviewers working for a market research firm specializing in mall–intercept interviews. On average the study took 15 min to complete. On completion, respondents were provided monetary compensation ($2) for their time, plus a bonus ($1) if they answered correctly a final quiz that tested their memory of the news story. Respondents were told about this bonus award at the start of the study to insure that the respondents read the stimulus materials carefully. Finally, participants were given a debriefing that explained the purpose of the study and uncovered the ‘cover story’ used in the manipulation.

3. **Results**

3.1. **Manipulation checks**

The manipulations worked as intended. Subjects in the negative condition rated OilCo’s actions toward the environment as more harmful (m=2.2; on a 1, harmful to 7, helpful scale), worse than other companies (m=2.1), and less socially responsible (m=2.1) than those in the positive condition (m=4.9, 5.3, 5.1, respectively; all p’s<0.001). As expected, there were no differences between the two groups in terms of perceived importance of company environmental responsibility, or attitudes toward buying from a company that was not environmentally responsible (both t’s<1.2, n.s.).

3.2. **Measures**

Two sets of questions tapped the ‘locus’ dimension. First, subjects were asked to rate the likelihood of various parties being a source of the engine lubricant problem (locus) on a 1–7-point scale (from not at all likely to very likely). They rated OilCo, retailers, consumers, and the supplier of the plastic container as potential sources of the problem. In a second set of locus questions, respondents were asked to assign a ‘percentage of the problem’ that might be due to each of the four parties (with totals summing to 100%). Locus was measured using the proportion of likely source judgments given to OilCo relative to all of the parties, and the percent given for the role played by OilCo relative to all of the parties. Both measures were proportions (r=0.48, p<0.001), and an average index was created from the two.
Stability was measured using an index comprised of four items: Do you think the problem represents something stable and ongoing with OilCo, or a fluke incident?, How likely is it that this type of problem will occur again in the future with OilCo products?, How likely is it that OilCo has had problems with its products in the past?, and How likely is it that the current problem with the engine lubricant is typical of OilCo products? (Cronbach’s \( \alpha =0.76 \)). Controllability was represented by a single item: How much control does OilCo have over this sort of product problem? Other, more specific, items used to measure controllability (e.g., How much control does OilCo have over the kind of plastic that is used in the jugs that they purchase from Carson Plastics), were significantly but insufficiently correlated with the more general item to allow the formation of an index with an acceptable \( \alpha \) level. Thus, the general item was used to tap controllability. (Alternative operationalizations of controllability, using the more specific items, were also analyzed and produced results similar to those reported below).

Three items comprised the index of blame: In your opinion, what is OilCo’s level of responsibility for the engine lubricant problem?, In your opinion, should OilCo be held accountable for the engine lubricant problem?, This incident is the fault of OilCo. Statements, such as the latter item, were answered on a 7-point totally disagree to totally agree scale (Cronbach’s \( \alpha =0.86 \)). Brand evaluations were measured by six items: Overall, what is your opinion of OilCo?, Generally speaking, what would you guess is the overall quality of the OilCo brand?, In your opinion, OilCo is not at all trustworthy/very trustworthy (7-point scale), …is not at all dependable/very dependable, …is not at all concerned about customers/very concerned about customers. Cronbach’s \( \alpha =0.92 \). Finally, two measures of buying intention were averaged: If you were shopping for an engine lubricant, how likely is it that you would by an OilCo lubricant? and If OilCo’s lubricants were priced 20% higher than a generic lubricant, how likely would you be to buy an OilCo lubricant? The correlation between the two items was \( r=0.78, p<0.001 \). All measures were coded such that higher numbers indicate greater perceived OilCo (internal) locus, stability, controllability, blame, positive brand evaluations, and purchase likelihood.

3.3. Tests of hypotheses

ANOVA’s are used to test our hypotheses that the corporate social responsibility manipulation affected attributions, judgments of blame, brand evaluation and buying intentions. Structural equation modeling results of our tests of mediation effects will then be reported.

As predicted by H1a, subjects in the positive CSR condition perceived the locus of the product crisis as more external (\( m=0.24 \)), while those in the negative CSR condition perceived locus to be more internal (\( m=0.36 \)). Control subjects (\( m=0.21 \)) gave ratings similar to positive condition subjects. [ANOVA for all three conditions, \( F(2,147)=13.09, p<0.001 \); contrast for positive vs. negative conditions, \( t(147)=3.95, p<0.001 \)].

Stability was also affected by the CSR manipulation, supporting H1b. Subjects in the positive CSR condition perceived the product failure to be less stable (\( m=4.1 \)) than those in the negative condition (\( m=4.7 \), with control subjects (\( m=3.9 \)) giving responses similar to those in the positive condition. [ANOVA for all three conditions, \( F(2,147)=4.38, p<0.05 \); contrast for positive vs. negative conditions, \( t(147)=2.23, p<0.05 \)].

The ANOVA for controllability across the three groups was less significant at \([ F(2,147)=2.64, p<0.08 \)]. There was not a significant difference between positive (\( m=5.4 \)) and negative condition subjects (\( m=5.8 \)), although the means differed in the expected direction \([ t(147)=0.99, n.s. \)]. Negative condition subjects did perceive greater controllability than did control group subjects \([ m=5.00; t(147)=2.29, p<0.05 \]).

Respondents in the negative condition blamed OilCo more for the incident than did those in the positive condition (\( m=5.5 \) and \( m=4.8 \), respectively) with control subjects (\( m=4.6 \)) making similar judgments to those in the positive condition. [ANOVA for all three conditions, \( F(2,147)=5.55, p<0.01 \); contrast for positive vs. negative conditions, \( t(147)=2.23, p<0.05 \)]. The social responsibility manipulation also affected brand evaluation and buying intentions. Subjects in the negative condition had a lower brand evaluation of OilCo (\( m=2.67 \)) than did those in the positive condition (\( m=4.25 \) with the control condition in between (\( m=3.73 \)). [ANOVA for all three conditions, \( F(2,147)=19.44, p<0.001 \); contrast for
positive vs. negative conditions, $t(147)=6.13$, $p<0.001$. Similarly, subjects in the negative condition had lower buying intentions ($m=1.94$) than did those in the positive condition ($m=3.89$) with the control condition in between ($m=2.77$). [ANOVA for all three conditions, $F(2,147)=20.47$, $p<0.001$; contrast for positive vs. negative conditions, $t(147)=6.37$, $p<0.001$].

Structural equation modeling (LISREL) was used to test the model and to estimate the standardized path coefficients shown in Fig. 1 (NNFI=0.97, CFI=0.99, RMSEA=0.067). Only those in the manipulated (positive and negative) conditions were included in the analyses ($n=100$). As predicted, the relationship between social responsibility and blame was mediated by the attributions made by subjects, supporting H2. The relationship between social responsibility and blame—using only the negative (coded 0) and positive conditions (coded 1)—was $-0.28$, $p<0.05$. When locus, stability, and controllability are included in the prediction of blame, this relationship vanishes ($B=-0.02$, n.s.). Social responsibility is a significant predictor of locus and stability ($B=-0.45$, $p<0.001$ and $-0.28$, $p<0.05$, respectively), but does not significantly predict controllability ($B=-0.14$, n.s.), although the effect is in the hypothesized direction. All three types of attribution are predictive of blame (locus: $B=0.37$, $p<0.001$; stability: $B=0.21$, $p<0.05$; and control: $B=0.22$, $p<0.05$). The degree of blame was consequential to brand evaluations ($B=-0.40$, $p<0.001$), even when controlling for the direct effects of CSR ($0.56$, $p<0.001$), as predicted by H3a. Consistent with H3b, brand evaluations predicted purchase intention ($B=0.83$, $p<0.001$). The model depicted in Fig. 1 accounts for 40% of the variance in blame, 60% of the variance in brand evaluation, and 65% of the variance in buying intentions. Two other models were also tested. The first model examines the direct and indirect effects of social responsibility on brand evaluation. The direct path from social responsibility to brand evaluation, without including blame (or attributions) in the model is 0.67, $p<0.001$. When blame is included in predicting brand evaluation this path drops to 0.56, $p<0.001$. Furthermore, the path from blame to brand evaluation ($B=-0.40$, $p<0.001$) is significant, as is the path from social responsibility to blame ($B=-0.28$, $p<0.01$). The second model examines the direct and indirect effects of social responsibility on buying intention. The direct path from social responsibility to buying intention is 0.72, $p<0.001$. When blame is included in predicting buying this path drops to 0.62, $p<0.001$. The path from blame to buying is significant ($B=-0.35$, $p<0.001$) and, as above, the path from social responsibility to blame is also significant ($B=-0.28$, $p<0.01$). Thus, social responsibility predicts brand evaluation and buying intention both directly and indirectly through blame.

4. Discussion

The results of this study provide support for the premise that CSR affects consumers’ attributions of
blame about product–harm crises. In particular, two of the three causal dimensions of attribution, locus and stability, were strongly affected by consumers’ prior perceptions of CSR, and controllability was affected to a lesser extent. In other words, consumers’ attributions about a product–harm crisis were found to be a function of consumers’ CSR associations. Each of the three attribution dimensions contributed significantly to consumers’ perceptions of blame for the product–harm crisis: internal locus, stability, and controllability were all positively related to blame. And, as expected, blame was found to have a strong negative effect on brand evaluations, which then predicted purchase intentions. The results indicate that CSR, as operationalized in this study, is a significant moderator of consumer attributions in a product–harm crisis, and that these attributions affect brand evaluations and purchase intentions. The implications of these results for CSR, for brand evaluations, and for attribution theory are elaborated on in the General discussion section.

In this first study, a fictitious company name was used because we were interested in the ability of our CSR manipulation to affect attributions of blame in a product–harm crisis. It has been suggested that the weak effects in previous research of the determinants of attributions were due to the use of realistic stimuli that dampened the effects of the intended manipulation (Folkes, 1988; Sparkman & Locander, 1980; Yalch & Yoshida, 1983). The use of a fictitious company in Study 1 eliminated the possibility of contamination of the manipulation by preexisting associations. Furthermore, given that the CSR manipulation constituted a large part of respondents’ corporate associations for OilCo, it could be argued that their effect on attributions is not surprising, and that demand effects may have been operating. In addition, this first study provides evidence of a main effect of CSR on attributions and brand evaluations. But as CSR research in routine consumer settings has shown, numerous variables moderate the impact of CSR on consumer behavior. Sen and Bhattacharya (2001), e.g., have shown that the impact of CSR is moderated by the extent to which consumers identify with the cause to which the company ascribes. We examine whether a similar moderating effect operates in nonroutine settings in influencing attributions. In particular, we build on research in psychology that has demonstrated that individuals tend to interpret event-related information and make attributions that are consistent with their prior beliefs (Crocker, 1980; Metalsky & Abramson, 1981). To examine the robustness of our findings, to eliminate the demand effects explanation, and to test for the effects of a moderating variable, we undertook a second study using a real company as the actor.

5. Study 2

5.1. Respondents, procedure, design, and measures

The purpose of Study 2 is to corroborate the essential results of Study 1 by eliminating alternative explanations, as well as to establish a boundary condition on the general effect of CSR on attributions uncovered in Study 1. Sen and Bhattacharya (2001) found that whether CSR influenced consumer evaluations or not depended on whether consumers saw congruity in the CSR domain they considered important and the one the company highlighted. We propose a more general form of this consumer-company congruence by positing that the effects of CSR will be more pronounced for those consumers who report that CSR is important to their decisions. The procedure for Study 2 was identical to that of Study 1. Again, 150 subjects participated. The average age of the sample was 36.3 (ranging from 17 to 83) and 61% were female. Ninety-seven percent reported having at least a high school diploma, with 46% of the sample reporting holding a college degree. Identical conditions and questionnaires to those used in Study 1 were utilized in Study 2, except that the fictitious company OilCo was replaced with a well-known real Oil Company. A measure of company familiarity was included in the questionnaire. Because respondents were familiar with the company (familiarity ratings averaged 5.15 on a 7-point scale) those in the control condition were also asked to evaluate the company on social responsibility. These measures permit corroboration of the impact of prior CSR, without the demand effects associated with the CSR manipulation. As in Study 1, respondents were debriefed at the end of the study, and were told that the information they had
read about the company’s environmental responsibility and products was developed for study purposes only and was fictitious.

6. Results

6.1. Manipulation checks

As in Study 1, the manipulations worked as intended. Subjects in the negative condition rated the company’s actions toward the environment as more harmful ($m=3.1$), worse than other companies (3.0), and less socially responsible (3.2) than those in the positive condition ($m=4.9, 5.1, 5.0$, respectively; all $p’s<0.001$). Means for the control group were closer to those of the positive condition (4.4, 4.5, and 4.8). There were no significant differences between the positive and negative groups (or the control group) in the perceived importance of company environmental responsibility, or attitudes toward buying from a company that was not environmentally responsible. Furthermore, there were no differences in familiarity with the company across conditions (all $F’s<2$, n.s.).

6.2. Measures

As in previous studies of the determinants of attribution, subjects were likely to bring preformed impressions of the company to the study. While these prior beliefs were clearly affected by the manipulations, it is unlikely that the social responsibility information we provided completely replaced prior impressions. More likely, the manipulations modified initial opinions. Thus, a continuous measure of CSR was created rather than using the manipulated variable. This allowed our analyses to reflect typical situations in which consumers’ prior impressions of a company are modified as new information is encountered through the press, advertising, product usage, etc. (and some consumers do not encounter this information, as represented by the control condition). Consequently, social responsibility was measured through the three manipulation check items (harm to the environment, treatment of the environment relative to other companies, and social responsibility; Cronbach’s $\alpha=0.90$). To parallel the analyses of Study 1, positive (greater than or equal to 5 on the scale, $n=65$) and negative (less than or equal to 3, $n=31$) prior CSR groups were created. All subjects, and thus the full range of social responsibility judgments, were included in the structural equation analyses presented below.

The two locus items were correlated 0.52, $p<0.001$, and the Cronbach’s $\alpha$ were 0.73, 0.81, 0.93 for stability, blame and brand evaluations, respectively. The correlation between the two buying items was 0.61, $p<0.001$.

6.3. Tests of hypotheses

Subjects with a positive prior CSR image perceived the locus of the product–harm crisis as more external ($m=0.21$), relative to those with a negative prior CSR ($m=0.34$), $[t(94)=3.94, p<0.001]$. The positive CSR subjects also perceived the problem as less stable ($m=3.7$) than the negative CSR subjects ($m=4.6$; $t(94)=3.95, p<0.001$). Judgments of control were not different between the two groups ($m=5.4$ and 5.2, respectively; $t(94)=0.60$, n.s.)$^4$. Thus, H1a and H1b were supported while H1c was not.

Those who held a negative corporate CSR image blamed the company more for the incident than those with a positive corporate CSR image ($m=5.3$ and $m=4.3$, respectively). The social responsibility manipulation also affected brand evaluation and buying intentions. Subjects with negative judgments of CSR had a lower brand evaluation (2.96) and lower buying intentions (2.02) than did those with positive judgments (m=5.50 and 4.04, respectively; $p’s<0.001$).

As in Study 1, structural equation modeling was used to test the model shown in Fig. 2, using all

---

$^4$ The same pattern of results is obtained if the manipulated conditions are used as the independent variable, rather than the measures of CSR. The only exception is that the difference in locus judgments across the positive and negative conditions is less significant ($p<0.07$). As in Study 1, the means for the control condition are close to those of the positive condition for locus and stability. For judgments of control, there are no significant differences across conditions.
subjects and the full range of social responsibility judgments (NNFI=0.92, CFI=0.97, RMSEA=0.10). As predicted, the relationship between social responsibility and blame was mediated by the attributions made by subjects, again supporting H2. The relationship between social responsibility and blame was $r = 0.29, p < 0.001$. When locus, stability, and controllability are included in the prediction of blame, this relationship drops to $r = 0.09$ (n.s.). Social responsibility is a significant predictor of locus and stability ($r = 0.36, p < 0.001$ and $r = 0.37, p < 0.001$, respectively), but does not significantly predict controllability ($r = 0.08$, n.s.), although the effect is in the hypothesized direction. Attributions of locus and control were significant predictors of blame ($r = 0.38, p < 0.001$ and $r = 0.28, p < 0.001$, respectively), while stability was nonsignificant but in the predicted direction ($r = 0.12$, n.s.). The degree of blame was consequential to brand evaluations ($r = 0.32, p < 0.001$), even when controlling for the direct effects of CSR (0.63, p<0.001), as predicted by H3a. Consistent with H3b, brand evaluations predicted purchase intention ($r = 0.67, p < 0.001$). The model depicted in Fig. 2 accounts for 39% of the variance in blame, 62% of the variance in brand evaluation, and 45% of the variance in buying intentions.

Two other models were also tested. The first model examines the direct and indirect effects of social responsibility on brand evaluation. The relationship between social responsibility and brand evaluation, without including blame (or attributions) in the model is 0.72, $p < 0.001$. When blame is included in predicting brand evaluation this path drops to 0.63, $p < 0.001$. Furthermore, the path from social responsibility to blame ($r = 0.29, p < 0.001$) is significant, as is the path from blame to brand evaluation ($r = -0.32, p < 0.001$). The second model examines the direct and indirect effects of social responsibility on buying intentions. The relationship between social responsibility and buying intention is $r = 0.50, p < 0.001$. When blame is included in predicting buying this path drops to 0.44, $p < 0.001$. The path from blame to buying is significant ($r = -0.19, p < 0.01$) and, as above, the path from social responsibility to blame is also significant ($r = -0.29, p < 0.001$). Thus, social responsibility predicts brand evaluation and buying intention both directly and indirectly through blame.

6.4. Moderation effects

In order to test for the predicted moderation effect, we split the sample into two groups based on their responses to the question asking how important it is that a company is environmentally responsible: the low importance group (less than or equal to 5, $n = 37$) and the high importance group (greater than 5,
Table 1

Moderation of importance of CSR

<table>
<thead>
<tr>
<th></th>
<th>Low importance of CSR</th>
<th>High importance of CSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple correlation</td>
<td>-0.04</td>
<td>-0.36***</td>
</tr>
<tr>
<td>CSR and Blame</td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ from CSR to locus</td>
<td>-0.02</td>
<td>-0.43***</td>
</tr>
<tr>
<td>γ from CSR to stability</td>
<td>-0.20</td>
<td>-0.41***</td>
</tr>
<tr>
<td>γ from CSR to control</td>
<td>0.21</td>
<td>-0.17</td>
</tr>
<tr>
<td>γ from CSR to blame</td>
<td>-0.04</td>
<td>-0.12</td>
</tr>
<tr>
<td>β from locus to blame</td>
<td>0.40**</td>
<td>0.37***</td>
</tr>
<tr>
<td>β from stability to blame</td>
<td>0.29*</td>
<td>0.08</td>
</tr>
<tr>
<td>β from control to blame</td>
<td>0.29*</td>
<td>0.25**</td>
</tr>
<tr>
<td>γ from CSR to brand</td>
<td>0.66***</td>
<td>0.62***</td>
</tr>
<tr>
<td>β from blame to brand</td>
<td>-0.38**</td>
<td>-0.33***</td>
</tr>
<tr>
<td>β from brand to buy</td>
<td>0.75***</td>
<td>0.66***</td>
</tr>
</tbody>
</table>

* p<0.05.
** p<0.01.
*** p<0.001.

n=112).5 Interestingly, those high in importance gave higher ratings for control (m=5.42) compared to those low in importance [m=4.51; t(148)=−3.24, p<0.01]. No other mean differences were found across the two levels of importance for any of the measures.

As predicted, for those who thought that a company’s environmental responsibility was unimportant, social responsibility did not predict blame, nor did CSR have a strong impact on attributions (See Table 1). The low importance group showed no simple relationship between CSR and blame (B=−0.04, n.s.) while the high importance group did (B=−0.36, p<0.001). This relationship remains nonsignificant for the low group (B=−0.04, n.s.) and drops to nonsignificant for the high group (B=−0.12, n.s.) when the attributions are added to CSR in the prediction of blame. Furthermore, CSR only predicts attributions for the high importance group (CSR to locus: B=−0.02, n.s. and B=−0.43, p<0.001, for the low and high groups respectively; CSR to stability: B=−0.20, n.s. and B=−0.41, p<0.001; CSR to control: B=0.21, n.s. and B=−0.17, n.s.). Attribution predicted blame in both models (locus to blame: B=0.40, p<0.01 and B=−0.37, p<0.001; stability to blame: B=0.29, p<0.05 and B=0.08, n.s.; control to blame: B=0.29, p<0.05 and B=−0.25, p<0.01). Thus, mediation only occurs in the high importance group. The rest of the model is similar between the two groups: the path from blame to brand evaluation (with CSR controlled) is B=−0.36, p<0.01 and B=−0.33, p<0.001, respectively for the low and high familiarity groups, the path from CSR to brand evaluations (with blame controlled) is B=0.67, p<0.001 and B=0.62, p<0.001, respectively, and the path from brand evaluations to buying intentions is B=0.75, p<0.001 and B=0.66, p<0.001, respectively.

6.5. Demand effects

To eliminate the demand effects explanation, we undertook an S.E.M. analysis of the data from only the control group subjects for whom CSR was not manipulated. For these subjects, measures of corporate social responsibility were taken at the start of the study, as they were not exposed to the manipulation of social responsibility. The mean rating of CSR was 4.66 (S.D.=1.13). Consistent with previous results, prior social responsibility predicts blame (B=−0.39, p<0.01), but drops to non significance when the three attributional mediators are included (B=−0.13, n.s.). Social responsibility predicts each of the dimensions of attribution: locus (B=−0.35, p<0.05), stability (B=−0.34, p<0.05), and controllability (B=−0.28, p<0.06). The coefficients for the attributional variables predicting blame are locus (B=0.14, n.s.), stability (B=0.33, p<0.05) and control (B=0.36, p<0.01). Blame predicts brand evaluations with the direct effects of CSR controlled (B=−0.22, p<0.05); and brand evaluations predict purchase intentions (B=0.58, p<0.001). Finally, prior judgments of CSR significantly predicted brand evaluation and buying intentions directly, and indirectly, through blame. Thus, this analysis replicates the findings for the full sample, and suggests that CSR effects do have significant impact on consumers’ attributions, brand evaluations, and purchase intentions, in the absence of demand effects due to the manipulation. Moreover, the finding of the moderating effects for the importance of environmental issues further reduces the possibility that demand

5 Most respondents gave high scores to this question. Thus, the sample was split this way to allow for a reasonable number of respondents in the low importance group. Given that 37 is a relatively small sample size for S.E.M., we confirmed that our results replicate using simple correlation and regression analysis for the high and low groups.
effects produced our results. Results due to demand effects tend to be straightforward and reflected in main effects. Moderating effects in between-subjects designs make demand effects far less likely because not only would subjects have to respond differentially to the demand characteristics of the study, those differential responses would have to be highly correlated with the moderating variable.

7. General discussion

Previous research has demonstrated the role of CSR in routine consumer behavior settings. The present research seeks to extend those findings by suggesting an alternative role for CSR in a nonroutine setting. The results show that CSR associations have a strong and direct impact on consumers’ attributions, which in turn translate into blame for the incident that consequently influences brand evaluations and purchase intentions. Furthermore, while CSR appears to influence brand evaluations directly, its impact through attributions appears to be pronounced only for those consumers who report considering a company’s CSR as important to their decisions. These results point to important theoretical and managerial implications.

That CSR associations in fact have a direct and relatively strong impact on consumer judgments in a nonproduct evaluation context suggests a new and interesting role for this construct. The findings suggest that CSR associations may have a significant impact when consumers rely on corporate associations to inform their judgments. The mediating role of attributions is important in these circumstances, because situations that are deemed out of the ordinary elicit spontaneous attributional activity. These attributions, we show, are strongly affected by consumers’ prior perceptions of CSR. In turn, these attributions contribute to an ascription of blame that affects brand evaluations, which then have an impact on purchase intentions. Thus, attributions and the consequent ascription of blame may have an enduring impact on consumer behavior by altering brand evaluations.

These results indicate that in addition to the effects of CSR on firm performance through improved consumer brand evaluations and greater likelihood of purchase in routine consumer behavior situations, CSR may have a ‘dormant’ effect that is activated in circumstances in which consumers rely on corporate associations to inform their judgments. Thus, our research suggests that even if positive CSR associations do not increase immediate profitability, they may be instrumental in reducing the risk of damage to brand evaluations in the event of a calamity. This permits a potentially novel conceptualization of the impact of CSR: CSR is like an insurance policy that is there when you need it.

Furthermore, the findings in Study 1 suggest a valence-based asymmetry in that a negative corporate CSR image has a larger impact on attributions than a positive CSR image. This suggests that a poor record on social responsibility can be asymmetrically damaging relative to the credit a firm receives for a good record. We find that a negative CSR image leads to unflattering attributions and blame while a positive image led to attributions similar to those made by control subjects who had no prior impression of the firm. One interpretation of this finding is that consumers are willing to give the benefit of the doubt to firms about which they know little, but that evidence of poor CSR places the firm in a pejorative position. The implication is that while a neutral image might provide as much protection in a product crisis as a positive image, a negative image will be a powerful liability to a firm facing such a crisis. The findings complement those of Dawar and Pillutla (2000) on product–harm crises, as well as those of Sen and Bhattacharya (2001) on CSR. Although, the Dawar and Pillutla (2000) research did not examine CSR, they found that companies about which consumers had weak prior expectations based on accumulated experience with the company, were barely able to maintain brand equity after a product–harm crisis, even if they responded positively and proactively to remedy the harm. Furthermore, if the weak prior expectation firm responded to the product–harm crisis by stonewalling, it suffered a disproportionately larger loss of brand equity than companies about which consumers had positive prior expectations. This finding corroborates the result in the Sen and Bhattacharya (2001) research in which all consumers were found to react to negative CSR information, but only those supportive of the CSR domain reacted to positive CSR information.
In addition to replicating the findings of Study 2 and ruling out a demand effect interpretation of our results, Study 2 also sheds light on a boundary condition of the effect of CSR. We find that CSR influences brand evaluations for all consumers, but that it only affects attributions when consumers see the CSR issue as important to them. This suggests that only those who care about the issue are motivated to access CSR information and to make attributions that are consistent with the firm’s CSR record. Moreover, it is only for these consumers that attributions will play a mediating role. These findings point to a contingency for the new role of CSR that the present research uncovers. Specifically, it appears that the ‘insurance’ role of CSR works primarily for consumers for whom CSR is an important decision criterion.

The results of this research also help sharpen our understanding of the role of attribution theory in consumer behavior. The findings show not just that differences in prior CSR result in differential blame for the product–harm crisis, but that this blame is in fact due to differences in consumers’ judgments of crisis locus, stability, and, to a lesser extent, controllability. These differences in the causal dimensions of attribution arise, despite exposure to identical crisis-related information. Two aspects are noteworthy. First, attributional judgments are malleable and depend on prior beliefs, and they mediate the relationship between prior beliefs (about CSR) and blame. In other words, the finding that judgments of an incident are more stable or more internal is a function of a firm’s prior CSR record demonstrates the subjectivity of attributions. Second, these judgments have a domino effect on brand evaluations and purchase intentions, suggesting that attributions can have enduring effects on consumer behavior, and are unlikely to be simply temporary artifacts of the product–harm crisis information.

7.1. Limitations and future research

The link between control judgments and blame was significant in both studies. CSR, however, did not act as a significant moderator of controllability, as it did for locus and stability. This finding may be due to a consumer perception that firms—as opposed to individuals, who were the focus of Weiner’s work—are expected to be in control of outcomes, whether they are socially responsible or not. This is supported by the relatively high ratings for control (between 5.2 and 5.8) given to both the positive and negative CSR firms across both studies. Thus, it appears that CSR plays its strongest role in judgments of whether the company continually runs into problems (stability) and whether it was the source (locus) of the crisis. Future research should examine whether there are circumstances (e.g., small or new firms) where control will play a greater role.

Our focus in this research is on the cognitive processes of attribution that are influenced by perceptions of CSR. Clearly, in a highly charged environment such as that of attributing blame for product–harm crises, emotional processes may also play a significant role. Other than measuring consumer evaluations, we have not examined how consumer emotions might affect or be affected by prior CSR perceptions. This remains a promising avenue for future research.

This research contributes to a better understanding of the role of CSR in consumer behavior. CSR effects on attributions and brand evaluations provide a promising avenue for further exploration. In particular, experimental research on consumers’ contingent use of CSR information can build on the moderation effects reported here to help define conditions under which CSR will or will not affect consumer behavior in nonroutine settings. Thus, this research has uncovered a novel role of CSR.

Acknowledgements

The authors are grateful to Pratima Bansal, June Cotte, Robert Fisher, and Zeynep Gurhan-Canli for helpful comments on earlier drafts. This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada and the Research and Development Department of INSEAD.

References


Organizational culture, market orientation, innovativeness, and firm performance: an international research odyssey

Rohit Deshpande\textsuperscript{a}, John U. Farley\textsuperscript{b,c,*}

\textsuperscript{a}Harvard Business School, Boston, MA 02163, USA
\textsuperscript{b}China-Europe International Business School, Shanghai, China
\textsuperscript{c}The Tuck School of Business at Dartmouth College, Hanover, NH 03755, USA

Received 14 March 2003; accepted 30 April 2003

Abstract

In this paper, we chronicle a research program spanning a decade and a dozen countries. The studies were framed in an extended model of competing values of organizational culture, and focused on how organizational culture, market orientation, and innovativeness affect the performance of firms competing in business-to-business markets. The design was developed and first tested in Japan. It was used throughout the research program, with the exception of a simplification of the sampling unit. We summarize substantive conclusions, including discussion of empirical similarities and differences in different cultures, in rich and poor countries, in the Orient and the Occident, in emerging economies, and in economies in transition towards a more market-driven form. As a general pattern, we find significant differences across countries in the means of all of the variables under study, and these differences generally reflect characteristics of national cultures. However, we find that slopes relating these variables to firm performance do not differ systematically over countries. We also examine buyer–seller relationships in the framework of an international “natural experiment” based on aspects of national culture. A brief research agenda is included.

Keywords: International management; Organizational culture and climate; Innovativeness; Market orientation; Firm performance; Business-to-business markets

1. Introduction

Deshpande and Webster (1989) proposed a research agenda for combining the study of several facets of organizational culture with the study of marketing management. In the first half of the 1990s, as discussed later, marketing also benefited from considerable progress in the measurement of market orientation and on relating these measures to firm performance. In 1993, a study of the effect of organizational cultures, market orientation (measured with a new scale), and innovativeness on the performance of Japanese firms was published (Deshpande, Farley, & Webster, 1993). The questionnaire was developed and administered in Japanese. As a general matter, hypotheses (discussed later) con-
ceming the relationship of performance to elements of the modified competing values model (also described later) were supported both quantitatively and qualitatively.

In the latter half of the same decade, the field of management in general (Management Science, 1994) and the field of marketing in particular (Marketing Science, 1995) turned attention to knowledge generalization—that is, to how existing knowledge can be used to predict what will happen in new, partially analogous product/market situations (Aharoni & Burton, 1994). The Marketing Science Institute (MSI) recognized the need for integration of organizational culture, market orientation, and innovation as linked to firm performance by designating interdisciplinary research leading to a better understanding of customer-oriented organizations as a highest research priority (Marketing Science Institute, 1994). Since this was also a period of intense interest in the rapid progress of globalization, there was particular interest in developing global generalizations of marketing knowledge (Winer, 1998).

Work expanded throughout the 1990s along the lines of testing a series of geographic generalizations of the research methodology and of substantive results developed in Japan. In the design of these extensions, the decision was made to maintain the same research methodology, retaining the advantages of direct comparability of measures, but at the same time losing the ability to really capitalize on the ongoing work in fields from which the research design was derived. Table 1 lists 9 published or forthcoming articles, which involved 10 countries (including pre-handover Hong Kong) and 17 cities. (To avoid highly repetitive citations in this paper, these articles and more recent working papers by the authors will be cited by the last initials and the year of publication—e.g. Deshpandé & Farley, 2002c; Deshpandé et al., 1993; other citations will be spelled out.) The articles are linked together by an effort to generalize conclusions, to the degree possible, quantitatively and non-contingently. The choice of countries and cities was largely driven by our research interests at the time, combined with some concerns with costs and the availability of skilled research suppliers.

The first step in the sequence was a test of generalizability in four other industrial countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Cities</th>
<th>Reported in</th>
</tr>
</thead>
<tbody>
<tr>
<td>First application a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>Tokyo</td>
<td>Deshpandé et al. (1993)</td>
</tr>
<tr>
<td>Generalizing in the industrial world a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>New York, Atlanta</td>
<td>Deshpandé et al. (2000)</td>
</tr>
<tr>
<td>France</td>
<td>Paris</td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>London</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>Frankfurt</td>
<td></td>
</tr>
<tr>
<td>Generalizing to the industrializing world a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>Bombay</td>
<td>Deshpandé and Farley (1999a)</td>
</tr>
<tr>
<td>Generalizing in Asia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td></td>
<td>Deshpandé and Farley (2001)</td>
</tr>
<tr>
<td>Vietnam</td>
<td>Hanoi, Ho Chi Minh City</td>
<td>Deshpandé and Farley (1999b)</td>
</tr>
<tr>
<td>Thailand</td>
<td>Bangkok</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Shanghai</td>
<td>Deshpandé and Farley (2000)</td>
</tr>
<tr>
<td>Generalizing to transition economies, including to state-owned enterprises</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>Hanoi, Ho Chi Minh City</td>
<td>Deshpandé and Farley (1999b)</td>
</tr>
<tr>
<td>China</td>
<td>Shanghai</td>
<td>Deshpandé and Farley (2000)</td>
</tr>
<tr>
<td>Generalizing to other cities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Beijing</td>
<td>Deshpandé and Farley (2002a)</td>
</tr>
<tr>
<td></td>
<td>Guangzhou</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shanghai</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shenzhen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tianjin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wuxi</td>
<td></td>
</tr>
</tbody>
</table>

* Quadrad design used.

(Deshpandé, Farley, & Webster, 2000), a near replication which is discussed in detail later in this paper to illustrate both the research methodology and the substantive conclusions. Later studies (chronological in completion but not necessarily in publication) focused on different issues. Second was a test, which
turned out to be successful, as to whether the methodology generalized to the industrializing world—in this case India (Deshpandé & Farley, 1999a). After a major simplification of the methodologies, two research streams followed. One stream involved Asian market economies (Hong Kong, India, and Thailand; Deshpandé & Farley, 2001) and the other focused on economies in transition from centralized economic planning to more market-driven structures (Vietnam, Deshpandé & Farley, 1999b, and China, Deshpandé & Farley, 2000). Finally, a comparison of six different cities was executed in culturally diverse China (Deshpandé & Farley, 2002a).

Later, two studies reanalyzed some of the data to examine the interplay of national cultures and buyer–seller relationships (Deshpandé & Farley, 2002a; SDF, 2000).

2. Conceptual structure: the elements of the modified competing values model

As noted earlier, the conceptual framework which links together this research program was motivated by a research agenda related to “organizational culture and marketing” proposed by Deshpandé and Webster (1989). This cross-national work has proceeded in a framework based on an integrative Theory of competing values of organizational cultures, which has been modified and expanded for use in the analysis of firm performance (Deshpandé et al., 1993). The modification involves incorporation of market orientation, innovation, and organizational climate into the structure.

2.1. Organizational culture and competing values

The starting point for the assembly of the modified model is the pattern of shared values and beliefs that help individuals understand how an organization functions (Deshpandé & Webster, 1989). Four classifications of culture were developed (Cameron & Freeman, 1991; Campbell, 1977; Quinn, 1988; Quinn & McGrath, 1985; Quinn & Rohrbach, 1983). Over the course of the work, we adopted new labels for the culture types, which are more closely related to the particular application. These new labels are used throughout this text; the original labels are shown in parentheses below:

- Competitive (market) culture characterized by an emphasis on competitive advantage and market superiority;
- Entrepreneurial (ad hocracy) culture emphasizing innovation and risk-taking;
- Bureaucratic (hierarchy) culture characterized by regulations and formal structures; and
- Consensual (clan) cultures emphasizing loyalty, tradition and internal maintenance.

In practice, our results indicate that organizations everywhere turn out to be a mixture of these four organizational culture types, but that the importance of each type differs significantly across countries in a pattern related to key attributes of national cultures. In terms of performance, the externally oriented cultures (entrepreneurial and competitive) should and generally do outperform the internally oriented (bureaucratic and consensual).

2.2. Motivation for modification of the competing values model—complexity in explaining performance

Both Deshpandé and Webster (1989) for conceptual reasons and Capon, Farley, and Hoenig (1996) for reasons related to a meta-analysis of studies of firm performance, suggested moving away from bivariate study (e.g. the effect of organizational culture on performance) and towards studies of related groups of determinants of performance—especially to groups of explanatory variables related to firm organization. The modification and expansion of the competing values model involved the addition of market orientation, innovativeness, and organizational climate. Each element of the resulting modified model of competing values has a history in terms of theory and measurement, and each was developed more or less independently by researchers in different fields—organizational behavior, marketing, and economic development.

2.2.1. Market orientation

The concept of market orientation is the central element of the management philosophy based on the marketing concept (Drucker, 1954; Levitt, 1960;
Menguc, 1996; Ruekert, 1992; Webster, 1988) and is presumed to contribute to long-term profitability. Because of the apparent importance of market orientation as a measure of successful implementation of the marketing concept, empirical work on the conceptualization, and measurement of market orientation was encouraged by the Marketing Science Institute in the mid-1980s (Deshpande, 1999; Deshpande & Farley, 1998, 1999c). Three different but related approaches were tested in the early 1990s to measure market orientation (Deshpande et al., 1993; Kohli & Jaworski, 1993; Narver & Slater, 1990). When used on the same subjects, the three scales were highly correlated and produced similar substantive results (Deshpande & Farley, 1998). The scales also consistently produced high reliability when used cross-culturally (DF, 1988).

The study of the market orientation–performance relationship in various industries and in various countries is very much an on-going research field. One branch of work involves the use of market orientation as a moderating variable in a more broadly specified model of firm performance (for example, Gatignon, Tushman, & Smith, 2002). Another line of development involves meta-analysis of the growing empirical literature. Work involving market orientation has also spread beyond the confines of marketing publications (for example, Appiah & Ranchhod, 1998).

2.2.3. Organizational innovativeness

Innovativeness, by hypothesis related positively to firm performance, has drawn interdisciplinary attention. The impact of innovation on firm performance and on economic growth has been of interest to economists for decades (Mansfield, Rapaport, Schnee, Wagner, & Hamburger, 1971; Schumpeter, 1934). Marketing has also been interested in innovation for some time. In one of his most-cited passages, Drucker (1954) linked innovativeness and market orientation, stating that “...business enterprise has two—and only two—functions: marketing and innovation...”. Innovation has been linked empirically to performance in the US (Capon, Farley, Hulbert, & Lei, 1991) and in China (Du & Farley, 2001). Because of the importance and the difficulty of managing innovation, students of organizations have been concerned with organizational designs which foster innovative behavior by managers (Gatignon, Tushman, & Smith, 2002).

2.2.3. Organizational climate

Organizational climate is an enduring quality of the internal environment of the firm which influences behavior and can be described in terms of attributes of an organization (Taguiri & Litwin, 1968). Empowered climates, which encourage communication, participation, decentralization, friendliness, and trust, have been related to performance in the US (Capon et al., 1996; Capon, Farley, Hulbert, & Lei, 1991). Organizational climate differs subtly from organizational culture due to a focus on the decision-making processes of the organization. Elements of organizational climate such as trust and work satisfaction have also been used as dependent variables in assessing such matters as the impact of downsizing and organizational restructuring (Lee, 2002).

2.3. Measuring performance

In the studies reviewed in this article, performance was measured with self-reported scales (Buzell & Gale, 1987) and validated for the (Deshpande et al., 2000) US data using publicly available historical data on the financial performance of the firms. Self-reporting is not an ideal solution to measuring performance, but it seems the most workable at the product/market level for making complex international comparisons. Different reporting requirements, ownership structures, and accounting practices pose great problems internationally in acquiring and comparing other sources of performance data. Even if these problems were absent, aggregate performance measures for a multi-divisional firm are often distant from the situation of a particular product/market situation. The performance scale used here has proven reliable in many different settings (DF, 1999c).

2.4. Each element of the modified model is a subject of ongoing research

Each of the four elements of the modified model is itself the subject of major on-going research development and has been over the past two deca-
des. This is shown in the numbers of published items relating each element to performance in peer-reviewed publications recognized by the ABI/INFORM database:

<table>
<thead>
<tr>
<th></th>
<th>Number of publications cited</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1986–1998</td>
</tr>
<tr>
<td>Organizational culture</td>
<td>96</td>
</tr>
<tr>
<td>and performance</td>
<td></td>
</tr>
<tr>
<td>Market orientation and</td>
<td>71</td>
</tr>
<tr>
<td>performance</td>
<td></td>
</tr>
<tr>
<td>Innovativeness and</td>
<td>33</td>
</tr>
<tr>
<td>performance</td>
<td></td>
</tr>
<tr>
<td>Organizational climate</td>
<td>48</td>
</tr>
<tr>
<td>and performance</td>
<td></td>
</tr>
</tbody>
</table>

Further, there are significant developments within each category which are not related explicitly to performance. For example, ABI cites 32 peer-reviewed publications between 1999 and the present related to organizational climate, and 202 more published between 1986 and 1998.

Work based on attempts to assemble more complex combinations of organization-related variables to explain aspects of performance, such as that described here, has grown more slowly. Only eight ABI-cited publications since 1999 involve intersections of three of the four elements of the modified model. The integrative contribution of our research is reflected in the fact that the only ABI-cited paper, which involved all four elements of the modified model was Deshpandé et al. (2000).

### 3. Method

The original methodology of the modified model was developed and tested in Japan to help establish the extent to which the elements of the model, which were generally developed in the West, would function in a non-Western setting. The original study also included development of a new market orientation scale because the other scales now available had not yet been published when data were collected. This new scale has been shown to be reliable in a wide variety of settings (DF, 1999c). The methodology reported in Deshpandé et al. (1993) formed the basis of the later studies after careful translation, appropriate modification of format, and extensive pre-testing in each new national setting.

#### 3.1. Samples

The studies focused on companies operating in business-to-business markets. This choice was made because it allowed direct identification of specific customers to participate by evaluating organizational cultures and market orientations for both themselves and a supplier. In the large industrial countries, as well as in India and in Hong Kong, the samples of firms selected for in-office personal interviews represent random samples of firms publicly headquartered in that country and in the city in which the interviews were conducted, and traded on the major stock exchange of that country. From this sample, a subsample of domestic firms headquartered in one major city was chosen, except for the US where the sample contained firms from two cities. (The geographic concentration was required because of the expense of the quadrad interviewing method described in Section 3.2.) In the transition economies, stock exchanges generally do not exist in most cities. Where they exist, they do not function like those in the market economies, so samples of locally based business-to-business firms in the transition economies were chosen randomly from available industrial directories. This was also done in Thailand to create a part of the sample consisting of members of the sizeable fraction of large firms which are privately (usually family) owned and not publicly traded.

#### 3.2. The initial unit of observation: the buyer–seller quadrad

Deshpandé et al. (1993) coined the term “quadrad” for the sampling unit involving double dyads made up of four interviews—two pairs each of matched buyers and sellers in a business-to-business relationship (Fig. 1). Besides the basic measures related to the competing values model, these provide measures of within-organization reliability (Moriarty & Bateson, 1982) and of buyer–seller agreement (Anderson & Narus, 1990; Siguaw, Simpson, & Baker, 1998; Weitz, 1981). Two marketing executives in a single business unit of each supplier firm were
interviewed in their offices about a specific product/market situation ("business"). The interviews were conducted in an appropriate local language by professional interviewers from a commercial market research firm. Two purchasing executives of a customer firm randomly chosen from a list of three such customers named by each supplier respondent were subsequently interviewed. These two respondents were chosen on the basis of their working with the supplier in question.

As Fig. 1 indicates, the design provides a very rich structure for the analysis of inter-rater reliabilities, validity of the supplier scales, and for new types of analysis, such as a gap which may exist between buyer and seller in the evaluation of the supplier’s market orientation. However, this sampling technique is cumbersome, and by economic necessity constrains both the geographic areas of the study because of travel cost. Also, the total number of observations is limited by the fact that each unit of observation (quadrad) requires four completed questionnaires.

The two relevant responses within the quadrad in each case were averaged to build the scales (Hage & Aiken, 1970), after finding no viewpoint variance problem (Heide & John, 1990).

3.2.1. The end of the quadrad

The research supplier in India reported difficulty in convincing both supplier firms and customer firms to provide the pair of required interviews, mainly because of relatively hierarchical structures. As a result, there were many sampled firms in which multiple respondents with both decision-making responsibility and with appropriate information were not available.

![Fig. 1. The quadrad design.](image-url)
This was not a problem in the relatively participative environments in the industrial countries that had been studied earlier. About the same time two reviewers and an associate editor observed that the quadrad structure had probably served its purpose, and as a result was probably no longer necessary in follow-up studies—advice that we were delighted to take in designing the work which followed the Indian study in the form of single respondents from supplier firms.

3.3. Measures, operationalizations, and questionnaire design

The sources of operationalizations and measurement content of the four explanatory constructs (organizational culture, market orientation, organizational climate, and innovativeness) as well as the performance measures are shown in Table 2. All scales were taken from other sources except for market orientation, which was developed and tested in Japan (Deshpande et al., 1993, which also includes an English version of the questionnaire in the appendix).

Our approach is intentionally etic, using comparative survey research methodology carefully adapted and pre-tested for each cultural environment (Deshpande & Webster, 1989; Moms, Leung, Ames, & Lickel, 1999). This etic measurement approach has well-known advantages and disadvantages (Berry, 1969, 1980). We hope that our results will complement related work using an emic approach, which is based primarily on ethnographic methods and which might be especially useful in this case in improving the organizational culture scales.

In the studies listed in Table 1, the scheme of Berry (1969) was used to establish elements of scale equivalence. Construct equivalence was tested in preliminary qualitative interviews with both senior managers and professional market researchers from each country to establish that the concepts served the same function from country to country (functional equivalence), that the concepts were expressed properly in each country (conceptual equivalence), and that the same classification scheme was usable in each country (category equivalence). Pre-tests of the questionnaire in each country ensured that scaling and measurement units were usable in that country (calibration equivalence). Two rounds of back-translation in each country established translation equivalence (Brislin, 1980). Sampling equivalence was provided by use of similar procedures and sampling frames in multiple countries.

4. Results

We do not attempt to catalogue details of the results from the studies listed in Table 1. Rather, we attempt to identify common elements from which we can draw generalizations, supplemented by some of
what we consider the more interesting specific findings. Section 4 contains the following:

1. Basic approach to analysis of cross-national differences on the measured values of the elements of the modified competing values model and analysis of country differences in how these elements affect performance.

2. Review of published comparisons in five industrial countries.

3. Discussion of consistencies in general patterns of results in six different settings.

4. Some of the more interesting specific results—for example, the degree of heterogeneity of cities in a culturally diverse country and potential usefulness of convenience samples.

5. Secondary research using part of the information to examine cross-cultural patterns of relationships in marketing.

4.1. Basic approach to analysis

The basic approach to analysis reflects the view of cross-national research of Farley and Lehmann (1994). They observe that easily detected, order-of-magnitude cross-national differences in average values of important variables create a sense of “everything is different” in the minds of marketing decision makers in international marketing. (Examples are large intercountry differences in per capita purchase of items like cameras (Armstrong, 1970) and in business spending per capita on advertising (Leff & Farley, 1980).) This mindset may carry over to an implicit assumption that the parameters of relationships between these same variables also vary a great deal internationally. In fact, these slopes often vary much less over countries or over time than do the means of the corresponding measures. This is consistently the case in the results in the comparative papers listed in Table 1.

The first step of the analysis is through MANOVA comparing and contrasting country means of the various measurements. The basic statistical approach to performance analysis is through segmented regressions, which contain specifications of country-specific intercepts and country-specific slopes for each explanatory variable. Later applications have added time-variant parameters to assess the impact of environmental shocks on system parameters (Deshpandé & Farley, 2002c).

4.2. An illustration: five industrial countries

To illustrate the approach, we examine the study carried out in five industrial countries (Deshpandé et al., 2000) and published in IJRM. Table 3 shows the values of the three elements of the analysis plan: the country means along with tests for mean inequalities, reliabilities of the scales, and regression results for pooled and unpooled models of performance.

---

Table 3
Country means of explanatory variables, scale reliabilities, and results of regression on performance

<table>
<thead>
<tr>
<th>Means</th>
<th>Reliability (Cronbach α)</th>
<th>β Coefficient in pooled performance regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japan</td>
<td>US</td>
</tr>
<tr>
<td>Organizational culture types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensual/clan</td>
<td>119.3</td>
<td>101.4</td>
</tr>
<tr>
<td>Bureaucracy/hierarchy</td>
<td>99.9</td>
<td>80.3</td>
</tr>
<tr>
<td>Entrepreneurial/ad hocracy</td>
<td>77.6</td>
<td>105.8</td>
</tr>
<tr>
<td>Competitive/market</td>
<td>103.3</td>
<td>112.5</td>
</tr>
<tr>
<td>Organizational climate</td>
<td>27.2</td>
<td>26.9</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>17.5</td>
<td>18.4</td>
</tr>
<tr>
<td>Market orientation</td>
<td>32.5</td>
<td>34.8</td>
</tr>
</tbody>
</table>

* Significant differences over countries (p < 0.01) and significant differences over culture types (p < 0.01) as a group in MANOVA.

* Significantly different over countries (p < 0.01).

* Regression coefficient significantly different from 0 (p < 0.05).

---

1 This section is adapted from DFW (2000).
4.2.1. Inter-country differences in scale averages

Based on MANOVA, the four organizational culture scores in Table 3 were significantly different as a group across countries. Market orientation is also significantly different across countries, with the US and Germany having the highest values and Japan the lowest. The Japanese sample is high on consensual culture, the US and England high on competitive culture, and France high on bureaucratic culture—all as expected, based on putative characteristics of national culture. Organizational climate and organizational innovativeness were not significantly different across these countries.

Pauly and Reich (1997) examine the loss of national characteristics in the face of globalization, concluding that there is only limited cultural convergence, which occurs in areas such as innovation management. Our results (which include US, Germany, and Japan which Pauly and Reich also studied) support this view. Organizational cultures are quite different and clearly relate to characteristics of national cultures. Market orientation also varies significantly, while innovativeness does not. The cross-sectional nature of our study does not allow us to assess movements towards or away from convergence.

4.2.2. Relating organizational factors to performance

Individual country and industry differences in the impact of the organizational variables on performance were tested using Chow (1970) tests on sub-elements of a segmented regression specified with country-and industry-specific slopes. The fit of the basic unsegmented model (all slopes constrained to be equal) was significant, but the inter-country mean and slope differences were not significant as a group. Country-by-country analysis showed that no individual country slopes were significantly different from the model where all intercepts and slopes were constrained to be equal for all countries. The covariate, industry, was not a significant factor in explaining performance.

Based on the lack of significance of specific country slopes or intercepts, a pooled model was used to examine the signs of the coefficients of a regression with performance as the dependent variable (last column of Table 3).

The organizational culture scores as a set have a significant effect on performance, and the magnitudes rank order as hypothesized in Deshpandé et al. (1993). Coefficients of competitive and entrepreneurial cultures are positive, with competitive being slightly higher. Coefficients of bureaucratic and consensual cultures are negative, with consensual the lowest. Participative climates that encourage communication, decentralization, and trust relate to better performance. Organizational innovativeness, the best-established of our independent variables in terms of its effect on performance, has the strongest single effect as is shown by the largest positive $\beta$ coefficient in Table 3. Market orientation had a positive sign but was not significant.

All $\beta$ coefficients have the expected sign, and a sign test rejects the hypothesis of a random sign pattern ($p < 0.01$). This indicates that substantive results on determinants of performance in Japan reported in Deshpandé et al. (1993) are not idiosyncratic to the special nature of that particular national context.

4.2.3. Scale reliability and invariance

Inter-rater reliabilities for individual items, measured by correlations of items from individual raters within the respective supplier and customer dyads, were statistically significant for all measures, with value correlation averaging around 0.65. Corresponding correlations were not significantly different for suppliers and customer dyads within individual quadrads.

All scale reliability estimates with scales pooled for the five countries (Cronbach $\alpha$) were at or above 0.6, and three were at or near 0.7.

All measures clear at least one criterion on invariance of covariance matrices of the different countries using the Steenkamp and Baumgartner (1998) procedure. The Box test found no significant country-specific differences of covariances of market orientation, organizational climate, or of three cultures—consensual, bureaucratic, and competitive. The consensual scale had partial scalar invariance. The entrepreneurial scale, the competitive scale, the market orientation scale, and the innovativeness scale all had some form of partial metric invariance.

4.3. Summary of performance results for six studies

Generalization of the Deshpandé et al. (1993) results were the focus of the five other studies listed in Table 1, all of which used a variant of the analysis structure described in Section 4.2. The study-by-study
summaries are shown in Table 4 and an overall summary is shown in Table 5. We assess the consistency of the results of these six studies in terms of presence or absence of mean differences of the seven organizational measures, and in terms of the hypothesized and actual relationship of the seven variables with performance. Also involved is attention to goodness of fit of the performance regressions, to scale reliability, and to invariance over countries of covariance matrices of the scales.

4.3.1. Intra-sample mean differences in scale averages

The general pattern of significant differences in the means of the organizational variables is consistent with those summarized for the five industrial countries discussed in Section 4.2. It is clear that there are differences in mean value of the seven explanatory variables, shown by the fact 23 of the 35 cases in which at least two populations are compared, have significantly different means. The means of each variable are significantly different in two or more of the five cases. Adhocracy culture is different in all five cases, market orientation in four cases, and organizational climate and innovativeness are different in two of five cases. All four organizational culture scores have different means in the case of the six Asian countries, which have highly diverse national cultures.

4.3.2. Slopes and fit in performance relationships

There was also a qualitative consistency of the partial relationships of the seven organizational variables with performance. Overall, 40 of the 41 regression coefficients have the hypothesized signs and 26 of these are significant. Innovativeness is positive and significant in all six cases. Market orientation is also positive in all cases and significant in four—especially outside the industrial world, where marketing is less developed and where investment in marketing may have higher pay-offs.

Organizational climate is not significant in the Asian countries, the transition economies, and the Chinese cities, while organizational climate has a significant effect in the industrial countries. There are indications that efforts of Western firms to install such Climate-related practices as entitlement in parts of Asia have not been effective, in part because of clashes with personal values (Ang, Lee, Singh, & Tan, 2000). There is a clear lack of significant effects (only one of eight of regression coefficients is significant) of organizational cultures in the Chinese cities and in the transition economies. On the other hand, both market orientation and innovativeness have significant effects on performance in the transition countries and in the Chinese cities.

The coefficient of determination of the performance regressions for the five industrial countries was 0.2, and for the six Asian settings 0.17, for transition economies 0.16, and for the Chinese cities 0.14. All were significant. Differences may be related to some extent to different between-study patterns in scale reliabilities discussed in the next section.

The magnitudes of the coefficients of determination are consistent with the contributions of a group of organizational structure and climate variables to explanation of performance of 121 large American corporations reported in Capon et al., 1996, p. 309. Coefficients of determination for fuller models, which also include measures of environment and strategy as well as organization structure and organizational climate, ranged from 0.41 to 0.57, depending on the measure of performance used.

4.3.3. Reliability and invariance

Generalization of the Deshpandé et al. (1993) results were the focus of five other studies, all of which used a variant of the analysis structure described in Section 4.2. We assess the consistency of the results of these six studies with those hypothesized (signal). We also compare goodness of fit of hypothesized relationships, reliability of scales, and invariance of covariance matrices of scales over countries.

4.3.3.1. Scale reliability. The first applications were in the industrial world, where pairs (dyads) of managers in supplier firms and customer firms established inter-rater reliabilities within firms, a result also confirmed in the quadrad analysis in India. Table 6 indicates that reliability declined as work moved outside the industrial world. Measures of market orientation, innovativeness, organizational climate, and performance had Cronbach’s above 0.6, and the majority had χ’s above 0.7. The majority of the scale reliabilities for organizational culture is above 0.6, and thus within a generally acceptable range for exploratory research of the type represented in this
Table 4
Summary for six studies of mean differences and relationships between performance and organizational variables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance relationship</td>
<td>Means different</td>
<td>Performance relationship</td>
<td>Performance relationship</td>
<td>Performance relationship</td>
<td>Performance relationship</td>
</tr>
<tr>
<td>Organizational culture (group)</td>
<td>*</td>
<td>NA</td>
<td>*</td>
<td>*</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Clan/consensual</td>
<td>- *</td>
<td>NA</td>
<td>-</td>
<td>yes</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Hierarchy/bureaucracy</td>
<td>-</td>
<td>NA</td>
<td>-</td>
<td>yes</td>
<td>-</td>
<td>no</td>
</tr>
<tr>
<td>Ad hocracy/entrepreneurial</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>yes</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Market/competitive</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>yes</td>
<td>+</td>
<td>no</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>yes</td>
<td>+</td>
<td>no</td>
</tr>
<tr>
<td>Organizational climate</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
<td>no</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Market orientation</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>yes</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Country (city) slopes different?</td>
<td>NA</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Industry significant?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>$R^2$</td>
<td>NA</td>
<td>0.20*</td>
<td>NA</td>
<td>0.17*</td>
<td>0.16*</td>
<td>0.14*</td>
</tr>
</tbody>
</table>

NA—not applicable.

* Indicates significance at $\alpha = 0.05$. 
paper. This pattern, which may be partially due to the fact that the four scales contain only four items, indicates that further research is needed on methodologies to measure organizational culture globally. Work on a more reliable scale to measure Innovativeness is also needed, as the items based on stages of the product life cycle used in these studies worked poorly outside the industrial world.

4.3.3.2. Invariance. During the later phases of the research described here, there was increasing concern with invariance of covariance matrices of the set of measures used to compare populations (generally meaning countries in international comparative research). Discussions often expressed concerns about US methods exported to the rest of the world, and we guess the same could be said about methods developed in Japan. Older methods such as the Box test of equality of covariance matrices can be used, and newer methods based on the measurement component of causal models (Singh, 1995; Steenkamp & Baumgartner, 1998; Steenkamp, ter Hofstede, & Wedel, 1999) are also available. These methods were applied to both the study of five industrial countries (DFW, 2000), where only the US and Japan had sample sizes adequate for testing, and to the six Asian economies (Deshpandé & Farley, 2001). Results were mixed. As mentioned earlier, all measures in the study of industrial countries clear at least one criterion of invariance. No sort of invariance was found for the six Asian economies, which also had much larger sample sizes.

4.3.4. Covariates—industry and firm size

Industry and firm size were used as indicators related to context of the studies.

Industry, divided into seven groups (financial and other services, consumer durables and non-durables, and industrial products, subdivided into capital goods, equipment, and supplies) was used as a covariate to represent possible differences in within-country and between-country business environments. In no case did the inclusion of Industry as a covariate have a significant effect on a coefficient of determination of a regression, nor did any have a significant correlation with any of the scales.

Firm size, when measured in terms of either revenues or number of employees, also was not correlated with performance or with any of the organizational scales.

4.4. A selection of more detailed results

The following sections include a number of what we consider interesting detailed results which occur in individual studies or subsets of studies. Some of these results are substantive and some are methodological, as is illustrated in Sections 4.4.1–4.4.4.

4.4.1. The transition economies

Chinese and Vietnamese firms have the highest scores on Bureaucracy of all countries studied, perhaps reflecting their development of many firms in an atmosphere of centralized economic planning. As was generally hoped with their adoption, the newer types of organizations (for example, joint ventures and wholly owned subsidiaries in China, and private firms in China and Vietnam) have lower bureaucracy scores and perform slightly better.

4.4.2. City effects

We are often asked whether focusing on a single large city in each country introduced bias into the results. Three studies examined differences between

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Summary for six studies of mean differences and relationships between performance and organizational variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organizational culture (group)</strong></td>
<td>Expected signs</td>
</tr>
<tr>
<td>Clan/consensual</td>
<td>–</td>
</tr>
<tr>
<td>Hierarchy/bureaucracy</td>
<td>–</td>
</tr>
<tr>
<td>Ad hocry/entrepreneurial</td>
<td>+</td>
</tr>
<tr>
<td>Market/competitive</td>
<td>+</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>+</td>
</tr>
<tr>
<td>Organizational climate</td>
<td>+</td>
</tr>
<tr>
<td>Market orientation</td>
<td>+</td>
</tr>
<tr>
<td>Country (city) slopes different?</td>
<td>NA</td>
</tr>
<tr>
<td>Industry significant?</td>
<td>NA</td>
</tr>
<tr>
<td>$R^2$</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA—not applicable.
cities within countries. Two cities were used in the US (New York and Atlanta (Deshpandé et al., 2000)) and in Vietnam (Hanoi and Ho Chi Minh City (Deshpandé & Farley, 1999b)) to identify the possible city differences in the values of the organizational variables or in the structure of the performance regression results. No significant differences were found in either.

The third study, designed at the strong urging of reviewers of earlier work, involved systematic study of city differences in six Chinese cities (Deshpandé & Farley, 2002a), where a number of factors might reasonably be expected to have shaped different organizational cultures. These cities have different spoken languages and millennia of different histories. Special economic zones, like the one in which Shenzhen grew from a village to a large city in ten years, might have made firms less like the profiles of cities undergoing economic transition more gradually, such as Beijing and Tianjin. Similarly, firms in cities more dependent on collective firms (Wuxi) may have a different set of cultures. Personal values are more traditional in firms in Beijing, Tianjin, and Wuxi than firms in Shanghai, Shenzhen, and Guangzhou, where the economic reforms had earlier impact. Again, we find city-specific differences in the means of some organizational variables, but we find that the slopes relating those measures to performance do not vary over six quite different PRC cities. The patterns of the mean differences are as might be expected. Firms in Shanghai, the engine of China’s economy as well as international trade up to World War II and now a leading center for industrial investment, show higher market orientation scores.

4.4.3. Convenience samples

Many studies of managers involve the use of convenience samples, leading to concern about representativeness of samples (for example, Aulakh & Kotabe, 1993). In Vietnam, we had the occasion to compare results from a sample of 100 firms drawn randomly with the results from a group of 27 managers from similar firms who attended an executive program together. Based on item-by-item t-tests, there were no differences in either mean or variance of any item in our questionnaire. Of course, this does not prove that convenience samples are representative, but it does show that such information can be quite useful.

4.4.4. The uniqueness of organizational culture in Hong Kong

Hong Kong provided the only sample (among 17) in which the pattern of the four organizational culture scores are statistically equal. This pattern may reflect the unique combination of Eastern and Western cultures, which has developed in Hong Kong over the past century. For example, Tung Chee-hwa said in his first public statement as the first Chinese leader of Hong Kong (New York Times, 1996), “We understand instinctively Chinese traditional values...yet we also appreciate...Western traits...”.

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Japan</th>
<th>Five industrial countries</th>
<th>Six Asian nations</th>
<th>China and Vietnam</th>
<th>Hong Kong</th>
<th>Six PRC cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensual/clan</td>
<td>0.42</td>
<td>0.65</td>
<td>0.63</td>
<td>0.47</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td>Bureaucratic/hierarchy</td>
<td>0.71</td>
<td>0.78</td>
<td>0.61</td>
<td>0.51</td>
<td>0.61</td>
<td>0.60</td>
</tr>
<tr>
<td>Entrepreneurial/adhocracy</td>
<td>0.66</td>
<td>0.60</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.51</td>
</tr>
<tr>
<td>Competitive/market</td>
<td>0.82</td>
<td>0.68</td>
<td>0.53</td>
<td>0.47</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>0.85</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.68</td>
</tr>
<tr>
<td>Organizational climate</td>
<td>NA</td>
<td>0.69</td>
<td>0.78</td>
<td>0.78</td>
<td>0.72</td>
<td>0.67</td>
</tr>
<tr>
<td>Market orientation</td>
<td>0.69</td>
<td>0.75</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.67</td>
</tr>
<tr>
<td>Performance</td>
<td>0.72</td>
<td>0.71</td>
<td>0.80</td>
<td>0.71</td>
<td>0.80</td>
<td>0.68</td>
</tr>
</tbody>
</table>
4.4.5. Other results from the quadrad

Despite the costliness and inherent difficulty of executing the quadrad, the design, in addition to inter-rater reliabilities discussed in Section 4.2.2, made possible a number of analyses, which could not have been done with either single informants or interviews of only supplier or customer dyads.

4.4.5.1. Comparing supplier and customer evaluations of supplier market orientation. The two measures of market orientation, one a self-evaluation and the other a customer evaluation, provide a good example of the distinction between differences in means and differences in slopes described in the basic approach to the analysis in Section 4.1. The self-evaluations are significantly larger on average than the customer evaluations, indicating that firms (as well as researchers) must be cautious in interpreting the average values of self-evaluations. On the other hand, the two market orientation measures are positively correlated, and regressions coefficients relating each to performance are not significantly different. Thus, the self-evaluations, which are generally easier to obtain, can be used to evaluate the impact of market orientation on performance.

4.4.5.2. Validity of organizational culture scales. Each customer provided their own organizational culture scores and scores for their supplier under study. The supplier’s self-evaluations and the customer’s evaluation of the supplier’s culture were not significantly different in the five industrial countries, providing a measure of validity for the organizational culture scales. The customer measures of their own culture were, as expected, significantly different from the supplier evaluations provided by the customers, again providing evidence of validity of the organizational culture scales.

4.4.5.3. Generalizing to the industrializing world. The quadrad design was used in a later study of Indian firms (Deshpandé & Farley, 1999a). Mean values of the scales were significantly different in India, where the scores on entrepreneurial culture were very high. However, regression coefficients of the organizational measures on performance were equal to those of other Asian countries (DF, 2001).

4.5. Secondary analysis of quadrad data: social identity theory and cross-cultural study of relationships

It is not unusual to return to a set of data to study matters not explicitly considered in the original study design. The quadrad allowed us to return for more detailed culture-related analysis of relationships between buyer and seller. Market orientation is based on the extent to which a seller’s focus on customers binds buyers and sellers together. Relationship marketing also focuses on the efforts of sellers, and to some extent buyers as well, to move away from an emphasis on transactions and towards investment in longer-term mutually profitable partnerships (Anderson & Weitz, 1989; Morgan & Hunt, 1994; Moriarty & Bateson, 1982). Secondary analysis of market orientation and relationship marketing used the social psychology framework of social identity theory for cross-national analysis of buyer–seller alliances (Deshpandé & Farley, 2001; Steinman, Deshpandé, & Farley, 2000).

4.5.1. The cross-cultural context of social identity theory: “us vs. them”

Elements of national culture have entered social identity theory (Tajfel & Turner, 1979) through the extent of inter-society differences in the importance of the individual and of the importance of groups in governing acceptable behavior in different cultures (Gannon, 2001; Kessing, 1974). Individualist cultures “emphasize the individual’s goals...individuals take care of themselves...”, while collectivist cultures “stress that group goals have precedence over the individual’s goals...the in-groups to which individuals belong take care of them in exchange for loyalty” (Gudykunst, 1989, pp. 166–167).

4.5.2. Linking market orientation and relationship marketing

Under social identity theory, marketers should try to get away from the “us vs. them” to the “we”—that is, to a sort of alliance based on consensus of buyer and seller. This is especially important in business-to-business markets where buyer and seller often engage in frequent direct personal contact. A conceptual basis for a connection between the two constructs is sug-
gested by Kalwani and Narayandas (1995), who stress looking at customer relations from a long-term perspective. Berry and Parasuraman (1991) also focus on customers: “Relationship marketing concerns attracting, developing, and retaining customer relationships”.

4.5.3. Measures: market orientation and relationship importance

The analysis involves relating market orientation as perceived and as desired to the importance of a relationship. We measure two types of market orientation: (1) actual market orientation (what the supplier respondents say about their own market orientation and what the customer respondents say about the supplier) and (2) normative market orientation (what suppliers and what customers think norms should be). The contents of the nine items, which comprise each scale, are shown in Table 2. Change in importance of relationship, a single item five-point scale measuring whether the relationship is becoming less or more important, has been used in research on industrial financial services (Bowman, Farley, & Schmittlein, 2000).

4.5.4. Results about gaps and relationships

Table 7 shows the size of the actual and normative market orientation gaps for low and high individualism countries as identified by the (Hofstede, 1980) individualism scales. Suppliers and customers do disagree about the supplier’s market orientation—that is, a market orientation gap does exist, and suppliers think better of themselves than customers do for both the actual and normative measures. The size of the gap increases as individualism of national culture increases. For Japan, the most collectivist culture, neither difference is significant, while the difference is significant for both the Middle countries (France and Germany) and for the Individualist countries (the US and UK). There is a negative correlation between both “gaps” and the scale measuring “relationship becoming more important”—that is, as either gap becomes larger, the relationship becomes less important.

Again, the results are not industry specific; in no case did the inclusion of industry as a covariate have a significant effect on the results.

As the normative and actual gaps decrease within a culture, relationship importance increases. For companies which base their marketing programs on relationship building, it could be important to assess these gaps and to narrow the specific areas of particular disagreement—as, for example, is indicated by individual items in the market orientation scales.

Closing the gaps discussed above to improve the “alliance” atmosphere will probably require changing perceptions of “us vs. them” to “we”—that is, away from the vocabulary of the transaction toward the

<table>
<thead>
<tr>
<th>Type of national culture</th>
<th>Country</th>
<th>Rank (of 39) on gap Hofstede (1980, 1991) individualism scale</th>
<th>Orientation gap (supplier’s self-evaluation minus customer evaluation)</th>
<th>t-test of differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collectivist</td>
<td>Japan</td>
<td>23</td>
<td>Actual 0.42</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normative 0.75</td>
<td>1.54</td>
</tr>
<tr>
<td>Middle</td>
<td>France</td>
<td>16</td>
<td>Actual 2.17</td>
<td>2.24*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normative 1.97</td>
<td>2.47*</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>11</td>
<td>Actual 3.50</td>
<td>3.23**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normative 2.90</td>
<td>2.01*</td>
</tr>
<tr>
<td>Individualist</td>
<td>UK</td>
<td>3</td>
<td>Actual 3.30</td>
<td>3.54**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normative 1.67</td>
<td>1.73*</td>
</tr>
<tr>
<td></td>
<td>US</td>
<td>1</td>
<td>Actual 2.73</td>
<td>3.69**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normative 2.61</td>
<td>5.16**</td>
</tr>
<tr>
<td>Correlation of “gap” with relationship becoming more important</td>
<td></td>
<td></td>
<td>Normative gap:</td>
<td>Actual gap:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( r = -0.16^* )</td>
<td>( r = -0.26^{**} )</td>
</tr>
</tbody>
</table>

Source: Deshpandé et al. (2000).
* Significant at \( p < 0.05 \).
** Significant at \( p < 0.01 \).
vocabulary of the relationship. Both social identity theory and our empirical results suggest that culture-related cross-national differences preclude benchmarking market orientation gaps with information from just one country.

5. Discussion

A framework built on an expanded theory of competing values was used to examine the impact of organizational culture, market orientation, organizational climate, and innovativeness on firm performance. These individual elements were developed more or less independently in different fields of research and at different times. The original design of the overall framework was developed in the early 1990s and used without major change for over a decade to provide comparability; during this period, there was significant research development in each of the four literatures which contributed research to the expanded competing values model.

The importance of the results is that improvement of a given size in a particular explanatory variable should be related to approximately the same improvement in performance in many countries. Open organizational cultures (competitive and entrepreneurial), stronger market orientation, and innovativeness all had a pattern of positive effects on performance as expected. Based on the insignificance of country effects in the slopes of regressions, we find similar substantive results in 10 countries, both industrial and industrializing, both Western and Asian, and in both market economies and economies in transition from central planning. Similar results hold qualitatively within a diverse set of Chinese cities. Further, the basic approach can be extended to analyses of such matters as the connection of relationship marketing with cultural differences.

5.1. Substantive generalizations

Some specific generalizations deal with both measures and markets:

- Market orientation and innovativeness have a pattern of consistently positive impact on performance. Innovativeness appears to be more important in the industrial world, and market orientation more important in the industrializing world, where the notion of marketing is at an earlier stage of development.
- As a general matter, relatively open, externally oriented organizational cultures related to better performance, while relatively closed, internally oriented organizational cultures related to poorer performance.
- In the so-called “transition economies” which are moving towards more market-driven economies, the results appear to apply to only a minority of firms—the upper third or upper quarter in terms of performance.
- Industry type, defined in rather broad categories (e.g. consumer non-durables), has no effect on these conclusions.

5.2. Methodological generalizations

The following generalizations have to do mainly with measurement:

- Market orientation, organizational climate, and performance were reliably measurable in all the situations studied. Organizational climate had limited impact on performance; it is possible that this measure is more specific to national culture than the others (Newman & Nollan, 1996).
- More specific issues can be examined by bringing new theories to bear on analysis of the same information. This was illustrated in the discussion of cultural patterns in buyer–seller relationships under social identity theory in Section 4.5.
- Work is needed on more universally reliable measures of organizational culture and innovativeness. The innovativeness scale is highly dependent on items related to the product life cycle, and these do not seem to work as well in the industrializing world. Similarly, the four-item organizational climate scales appear too limited in the industrializing world.
- In terms of methodology, the exact role of invariance of individual country covariance matrices is unclear, as are the sensitivities of invariance tests to sample sizes, scale reliability, collinearity of measures, or sensitivity of the tests to various types of alternatives. Probably more important is
developing a better understanding of the effect on both reliability and invariance of relatively high levels of observation-idiosyncratic variability in the dependent and independent variables, which apparently occur in cross-sectional studies of firm performance. Experiments, perhaps in a simulation framework, might be as useful in this situation as they were in the early history of econometrics in assessing the actual effect of specification error on various approaches to parameter estimation. Invariance will probably continue to be a nearly ubiquitous problem for cross-national researchers. Steenkamp et al. (1999) suggest that using alternative tests is useful, but that estimation can produce meaningful results in absence of invariance. (Ryan, Chan, Ployhart, & Slade, 1999), in a rather balanced discussion of the subject, argue that invariance should at least be considered, but that some methods appear to be too stringent. They also point out that a model that predicts results well in each country (that is, that has a consistent “signal” in terms of parameter estimates and model fit) may help override some apparent weaknesses in data.

5.3. What next?

The results, while consistent with theory and consistent over countries, probably raise more questions than provide definitive answers. Some questions deal with limitations on generalization:

- These studies, while they use representative samples of manufacturing and service firms, deal only with direct business-to-business relationships. Studies of patterns of perceptions of market orientation through the layers of complex distribution systems, for example, in the business-to-consumer context, would broaden the scope of generalization.
- The growing interest in causal models during the period of the research program suggests returning to the analysis in this context. (Throughout this research program, the scales were maintained in the form in which they were developed—that is, with items weighted equally.) As a next step, confirmatory factor analysis should be explored as a means to weight the elements of the various scales, perhaps leading to the addition of a measurement model to future studies of this type.
- The quadrad design, which required limiting interviewing to geographically proximate customers, left open the question of whether geographic distance between customer and supplier is systematically related to perceived levels of market orientation.
- Most of the results deal with a single, major city in each country. The results on the existence of within-country city patterns are similar rather than different.
- All individual studies are cross-sectional. Longitudinal studies would be useful to study changes over time in perceptions of managers of both producing and buying firms. A useful design might include periodic re-interviews on a subset of a sample in panel form, with the rest of each sample being used in fresh samples to help calibrate any effects of re-measurement (Deshpandé & Farley, 2002c).
- The choice of countries was driven more by chance than by design. In particular, a balanced approach to characteristics of the sample countries—culture, demographics, major institutional factors, etc.—could help sharpen results.
- In all likelihood, industry (which did not effect our results) probably does have a systematic effect at some level of greater disaggregation. Many students of firm performance find that industry characteristics systematically matter in study of performance (Bain, 1951; Hansen & Wernerfelt, 1989; Scherer & Ross, 1990), hence probably in organizational elements also related to performance, such as market orientation. Some experiments with different levels of aggregation might be useful.

Some open questions deal with methodological matters:

- As mentioned earlier, some basic work is needed on the scales measuring organizational culture and innovativeness.
- Experiments with measures of performance other than self-reports would be most useful, although self-reports have proven reliable in many settings.
- Replication and re-measurement should be considered to track stability of results over time and to assess systematic changes related to major envi-
5.4. Towards a broader perspective on determinants of firm performance

While the results do seem robust, our research program is based on three classes of essentially organizational variables. This helps remedy the relative historical void of organizational measurements (at least relative to environmental and strategic measures) in studies of performance. However, it is important that these results now be fit into a broader context in terms of a wider range of determinants of firm performance. The research required to develop a broad understanding of superior firm performance is daunting. We know a great deal about the bivariate effect on firm performance of a long list of environmental and strategic factors, but we know relatively little about how these elements combine within categories and even less about how environmental, strategic, and organizational combine across categories in a comprehensive model of firm performance.

Acknowledgements

The authors are indebted for support of these projects to the China–Europe International Business School, the Center for Asia and the Emerging Economies of the Tuck School, and the Faculty Research Fund of the Harvard Business School.

References


Customer satisfaction and loyalty in online and offline environments

Venkatesh Shankara,*, Amy K. Smithb,1, Arvind Rangaswamyc,2

a Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA
b George Washington University, School of Business and Public Management, Washington, DC 20052, USA
c The Smeal College of Business, Penn State University, University Park, PA 16802, USA

Received 14 November 2000; received in revised form 8 November 2002; accepted 14 November 2002

Abstract

We address the following questions that are becoming increasingly important to managers in service industries: Are the levels of customer satisfaction and loyalty for the same service different when customers choose the service online versus offline? If yes, what factors might explain these differences? How is the relationship between customer satisfaction and loyalty in the online environment different from that in the offline environment? We propose a conceptual framework and develop hypotheses about the effects of the online medium on customer satisfaction and loyalty and on the relationships between satisfaction and loyalty. We test the hypotheses through a simultaneous equation model using two data sets of online and offline customers of the lodging industry. The results are somewhat counterintuitive in that they show that whereas the levels of customer satisfaction for a service chosen online is the same as when it is chosen offline, loyalty to the service provider is higher when the service is chosen online than offline. We also find that loyalty and satisfaction have a reciprocal relationship such that each positively reinforces the other, and this relationship between overall satisfaction and loyalty is further strengthened online.

Keywords: Online markets; Satisfaction; Loyalty; Services, Internet; E-Commerce

1. Introduction

The rapid growth of online transactions in service industries raises important research questions about the levels of satisfaction and loyalty in the online environment, and the relationship between satisfaction and loyalty online relative to offline. Compared to the offline environment, the online environment offers more opportunities for interactive and personalized marketing (Wind and Rangaswamy, 2001). These opportunities may influence customer satisfaction and loyalty differently in the online environment vis-à-vis the offline environment.

Managers are concerned about how the online medium influences satisfaction and loyalty and the relationship between satisfaction and loyalty. Typically, online customers can more easily compare...
alternatives than offline customers, especially for functional products and services. A competing offer is just a few clicks away on the Internet. Because of these properties of the Web, many managers fear that the online medium may induce lower customer satisfaction and loyalty compared to the offline medium, and that increased satisfaction with a service may not lead to higher loyalty when that service is chosen online.

Some recent studies show that there may be systematic differences in customer attitudes and behavior for products and services chosen online versus offline. For example, price sensitivity may actually be lower online than offline (e.g., Degeratu, Rangaswamy, & Wu, 2000; Lynch & Ariely, 2000; Shankar, Rangaswamy, & Pusateri, 2001). Brand names (brand equity) could also have higher impact online than offline (Degeratu et al., 2000). Collectively, these studies suggest that a priori, customer satisfaction and loyalty for services may be different online (compared to offline), and not necessarily lower online. However, we do not know of any academic research that has carefully explored whether the online medium positively or negatively impacts customer satisfaction and loyalty (other things equal) and why.

Satisfaction and loyalty are not surrogates for each other (Bloemer & Kasper, 1995; Oliver, 1999). It is possible for a customer to be loyal without being highly satisfied (e.g., when there are few other choices) and to be highly satisfied and yet not be loyal (e.g., when many alternatives are available). Firms need to gain a better understanding of the relationship between satisfaction and loyalty in the online environment to allocate their online marketing efforts between satisfaction initiatives and loyalty programs. If, for example, the firm finds that loyalty is associated with increased satisfaction, it could directly focus on enhancing its loyalty programs.

In this paper, we address the following questions: (1) For a given service, other things equal, are the levels of customer satisfaction and loyalty different when the service is chosen online as compared to offline? If yes, what factors might explain these differences? (2) Is the relationship between satisfaction and loyalty stronger or weaker online compared to the offline environment? To answer these questions, we develop a set of hypotheses based on a conceptual framework. To test the hypotheses, we formulate simultaneous equation models and estimate them using data from two sets of online and offline samples in the lodging sector of the travel industry. An important characteristic of industries such as online travel, entertainment (e.g., movie, theater, and concert tickets), and restaurant is that regardless of whether customers choose the service provider online or offline, the actual service is experienced by the customer offline. This characteristic enables us to isolate the effects of the medium on satisfaction, separate from the effects of service attributes.

Prior research studies on satisfaction and loyalty have primarily been conducted in the offline environment. We extend previous research in many ways. First, to our knowledge, ours is the first empirical study to compare the effects of the medium (online versus offline) on customer satisfaction, loyalty, and the relationship between satisfaction and loyalty. Second, we study the reciprocal relationship between customer satisfaction and loyalty that offers the potential for deeper insights into the nature of the relationship between these constructs. We also empirically test this reciprocal relationship. Much prior research has focused on the impact of customer satisfaction on loyalty, but not vice-versa. Third, unlike previous studies that have generally focused on action/behavioral loyalty, we focus explicitly on attitudinal loyalty to the service provider. Attitudinal loyalty is similar to affective/conative loyalty proposed by Oliver (1999) and represents a higher-order, or long-term, commitment of a customer to the organization, which cannot be inferred by merely observing customer repeat purchase behavior. Customer retention can occur without attitudinal loyalty, if for example, the customers are indifferent, or there are no other viable choices in the market.

2. Conceptual framework and research hypotheses

Consistent with Oliver (1999), we define satisfaction as the perception of pleasurable fulfillment of a service, and loyalty as deep commitment to the service provider. We focus on attitudinal loyalty, rather than on behavioral loyalty (Day, 1969; Dick & Basu, 1994) for the following reasons. A behaviorally loyal customer may be spuriously loyal, that is,
stay with an organization or service provider until he/she can find some better alternative in the marketplace (Dick & Basu, 1994). An attitudinally loyal customer, on the other hand, has some attachment or commitment to the organization and is not easily swayed by a slightly more attractive alternative. Attitudinal loyalty not only indicates higher repurchase intent, but also resistance to counter-persuasion, resistance to adverse expert opinion, willingness to pay a price premium, and willingness to recommend the service provider to others.

We selected the travel industry as the context for our study for three reasons: (1) Although people can make travel choices (reservations) online, they still have to experience the service offline. Therefore, the actual service encounter itself is no different if the same service provider and service options are chosen online or offline. (2) The travel industry is one of the largest industries online and has had a relatively long history of online presence. Thus, many people are comfortable making travel choices online. (3) People make travel choices both online and offline, as compared to a pure online service provider like Yahoo, which has no direct offline equivalent. This allows us to compare online satisfaction and loyalty with offline satisfaction and loyalty for the same service.

We develop hypotheses about the effects of constructs focal to our research, namely, the direct effect of the online medium on service encounter satisfaction, overall satisfaction, and loyalty. We also develop hypotheses about the relationship between satisfaction and loyalty and the differential effects of various factors such as ease of obtaining information and frequency of use, online relative to offline. In developing the hypotheses, we primarily draw upon behavioral decision theory that examines tradeoffs between cognitive effort deployed and the quality or accuracy of the decisions (choices) that consumers make (Johnson & Payne, 1985; Johnson, Bellman, & Lohse, 2002). We also predict the effects of control variables such as website factors, service attributes, and prior experience with the service, on satisfaction and loyalty.

Fig. 1 summarizes the conceptual model of how the online medium influences customer satisfaction and loyalty. The figure includes the focal constructs as well as control variables. Our main thesis is that satisfaction, loyalty and their relationship differ online.
from offline because the same customers could make different choices online versus offline in the same decision situation. We now describe each of the constructs used in the study.

2.1. Drivers of customer satisfaction

We examine two types of customer satisfaction, namely, service encounter satisfaction and overall customer satisfaction. Service encounter satisfaction is transaction-specific, whereas overall customer satisfaction is relationship-specific, that is, overall satisfaction is the cumulative effect of a set of discrete service encounters or transactions with the service provider over a period of time (Bitner & Hubbert, 1994; Oliver, 1997; Rust & Oliver, 1994). Although these two types of satisfaction are related, it is important to recognize them as distinct constructs because some of the factors influencing them may be different. For example, service encounter satisfaction is more likely to depend on performance on specific attributes of the service encounter (e.g., Was the price consistent with expectation?), whereas overall satisfaction is more likely to depend on factors that occur across transactions (e.g., Is it easy to shop for this service?).

Table 1 summarizes several previous studies of the determinants of customer satisfaction. Based on these studies, we expect that customer satisfaction may be influenced by service provider performance on various service attributes (e.g., location of a hotel and its reputation), customers’ prior experience with the service, their frequency of service use, disconfirmation of time spent choosing a service (e.g., whether it took longer or shorter than expected to make a choice), and ease of obtaining information about the service. We also argue, for reasons discussed below, that loyalty to the service provider also impacts customer satisfaction. These factors should apply to both online and offline environments. Below, we elaborate on each of the factors affecting customer satisfaction.

2.1.1. Focal constructs

2.1.1.1. Direct (main) effect of the online medium (+). The online medium has several distinctive features that may directly impact both service encounter satisfaction and overall satisfaction. There are several reasons why the online medium influences customer satisfaction in a different way than the offline medium.

- Similar to other self-service technologies (SST), the online medium could improve the shopping process by enabling customers to sort (e.g., sort hotel rooms by price, miles from nearest airport, etc.) and group information (e.g., categorize hotels by quality rating), by increasing the number of options available (e.g., number of different hotels available near a given location), and by enabling customers to access peer opinions and ratings (Meuter, Ostrom, Roundtree, & Bitner, 2000). Thus it is likely that, for the same customer, online and offline decisions take place under different sets of information. Potentially, there is more information available online. With more available information, customers are likely to devote more cognitive effort to their decision processes because they can see the potential for realizing addi-
tional benefits (e.g., more informed or improved choices; lower prices) through additional effort (John-
son & Payne, 1985). Additional information (possibly in conjunction with more effort) will improve the
quality of the choices that customers make, which in turn, is likely to result in a service experience that
delivers higher satisfaction when the choices are made online than offline.

- The online shopping experience could also alter customer expectations about the services—if they
have more information (including visual information) beforehand, they know what type of service they will
get and are less likely to be surprised or angry at the service received, than when they make a choice off-
line. This suggests less disconfirmation with expect-
ations (i.e., lower positive or negative disconfirma-
tion), which should also result in greater satisfaction than when the choice is made offline.

- It is also possible that the online environment
could decrease satisfaction because of the perceived lack of privacy and financial security, perceived lack
of human contact, failure of technology, and poor design of interface (Meuter et al., 2000).

The direct effect of the online medium on customer satisfaction would depend on the net effect of these
sources of satisfaction and dissatisfaction and is an empirical issue. However, we offer the following
hypotheses a priori as our best judgments of the directionality of the combined effects of the medium
on satisfaction.

H1a: Customers who choose a service online are more
likely to have a higher level of service encounter sa-
tisfaction than those who choose offline.

H1b: Customers who choose a service online are more
likely to have a higher level of overall satisfaction
than those who choose offline.

2.1.1.2. Loyalty (+). A loyal customer is more likely
to find the service encounter and the overall experi-
ence with a service provider more satisfying than a
nonloyal customer. Although previous research has
not viewed loyalty as an explanatory variable of satisfaction, prospect theory (Kahneman & Tversky,
1979) offers a rationale for why loyal customers may
be more satisfied than nonloyal customers (see also,
Mittal, Ross, & Baladasare, 1998). According to this
theory, losses loom larger than gains. In the context of
the relationship between loyalty and satisfaction, this
theory suggests that if loyal customers have a negative experience and become dissatisfied with a service
provider, then they might gain by switching to a new provider. However, by switching they will incur
loss in the form of loyalty benefits (e.g., emotional loss, rewards program) or face a potentially unfamiliar
service encounter. Compared to nonloyal customers, loyal customers may perceive the loss to be larger
than the short-term gain of moving to a new service provider. Research reported by Ahluwalia, Unnava,
and Burnkrant (1999) provides further support for the loyalty—satisfaction relationship. They find that loyal
customers are much less susceptible to negative information about a product than are nonloyal customers.

With respect to online customers, they recognize
that they have chosen their favored service provider
even when confronted with a large number of other
options that were just a click away. Thus, they may feel partly responsible if the chosen service does not
fully live up to their expectations, thereby mitigating the impact of dissatisfying experiences. Finally, some
website features (e.g., information on related services,
better use of rewards, and personalization) allow loyal
online shoppers to derive greater utility and be more satisfied, as compared to shopping through the offline
medium which offers limited pre-consumption inter-
actions with the service provider. Taken together,
these arguments lead to the following hypotheses.

H2a: The positive effect of loyalty on service en-
counter satisfaction with that service provider is
greater for customers who choose online than it is for
those who choose offline.

H2b: The positive effect of loyalty on overall
satisfaction with that service provider is greater for
customers who choose online than it is for those who
choose offline.

2.1.1.3. Ease of obtaining information (+). Overall
customer satisfaction may also depend on the ease
with which information relevant to a particular cus-
tomer can be obtained while choosing the service
provider (Oliva, Oliver, & MacMillan, 1992). With
more relevant information, customers make better
decisions leading to higher satisfaction. Easier access
to information also typically increases customer sat-
satisfaction with the shopping process (unless there is
information overload), which could increase overall customer satisfaction with the service—i.e., there could be some carryover of (dis)satisfaction with the shopping process to (dis)satisfaction with the service provider.

The online medium could further enhance the positive effect of information access on overall satisfaction by making the information obtained more actionable in influencing choice (for example, the information could contain direct links to specific hotel properties), thus involving less cognitive effort online than offline. Also, for such services as hotels, which have a large number of search attributes (e.g., location, price, type of bed), it is easier to obtain information online than offline. For comparable levels of information access and use, the lower effort deployed online is likely to lead to greater satisfaction online than offline (Johnson & Payne, 1985; Meuter et al., 2000). These arguments suggest the following hypothesis.

**H3:** The positive effect of ease of obtaining information on overall customer satisfaction for a service provider is higher for customers who choose the service online than it is for those who choose offline.

2.1.1.4. Frequency of service use (+/-). Usage frequency may impact both service encounter satisfaction and overall satisfaction (Bolton & Lemon, 1999; Vredenburg & Wee, 1986). Arguments can be made for both a positive and a negative relationship between usage frequency and satisfaction. Supporting a positive relationship is the argument that if customers use a service often, the repeated exposure to favorable service encounters may continually lead to higher satisfaction. Supporting a negative relationship is the concept of expectation-performance. When customers use a service often, they start to treat the benefits of the service as a given and expect additional benefits from the service provider overall, and in each service encounter. The raised expectations may lead to lower satisfaction. Prior studies have more commonly reported negative effects than positive effects.

The online medium could dampen this negative effect of usage frequency on overall satisfaction. Because the online medium generates expectations that are more consistent with the actual service levels (e.g., the availability of room views and other relevant information while choosing a hotel), it mitigates problems associated with frequent users having higher expectations, and therefore, being potentially less satisfied with the service provider. These consistent expectations also lower the mental costs of online choices, improving both utility and satisfaction (Johnson & Payne, 1985). These arguments lead to the following hypothesis.

**H4:** The negative effect of frequency of use on overall customer satisfaction for a service provider is lower for customers who choose the service online than it is for those who choose offline.

2.1.2. Control variables

2.1.2.1. Service attribute performance (+/-). The actual performance on various service attributes that a customer experiences during a service encounter influences customer satisfaction with that service (Bearden & Teel, 1983; Bolton & Drew, 1991; Mittal et al., 1998; Oliva et al., 1992; Oliver, 1993; Spreng, MacKenzie, & Olshavsky, 1996). Because the actual service levels may vary from one service encounter to another, we do not expect transaction-specific service attribute ratings to directly influence overall customer satisfaction, but we expect them to have an indirect effect on overall satisfaction through service encounter satisfaction. Attribute-level performance can influence satisfaction either positively or negatively. For example, the better the room amenities relative to customers’ expectations the higher their satisfaction, whereas the higher the price relative to customers’ expectations the lower their satisfaction.

2.1.2.2. Prior experience (+). We expect a customer’s prior experiences with a service provider to strongly influence both satisfaction with the service encounter and overall satisfaction. Woodruff, Cadotte, and Jenkins (1983) argue that a favorable prior experience with a service provider increases the likelihood of a favorable evaluation of the current service encounter as well as the overall evaluation of the service provider by affecting their norms and expectations. They showed that customers’ past experiences with restaurants affect their evaluations of subsequent dining experiences. In a study of the auto industry, Vredenburg and Wee (1986) found that favorable
prior experience resulted in higher satisfaction levels. Bolton and Drew (1991) showed that customers’ prior attitudes influenced their current attitudes toward a telephone service. Thus, overall, the more favorable the prior experience the higher the satisfaction. Unlike Frequency of service use, which refers to the quantity of prior use of a service, Prior experience refers to the quality of previous experiences with the service provider.

2.1.2.3. Organizational constraint on choice (−). Service encounter satisfaction is likely to depend on whether the choice of the service on a particular occasion was restricted by organizational constraints. For example, customer satisfaction during a service encounter would be more positive if he/she chose the service provider than if his/her organization chose it.

2.1.2.4. Role of intermediary in choice (?). If an individual chooses a service after going through an intermediary such as a travel agent, he/she is likely to have different expectations about that service than if he/she chooses without that intermediary for that service encounter. The expectations could be higher or lower depending upon the influence of the intermediary. The different expectations will likely lead to a different assessment of satisfaction for the service encounter.

2.1.2.5. Service package (?). An individual’s service encounter satisfaction with a service provider (e.g., hotel) may depend on whether the choice of the service was part of a package or bundle of services (e.g., travel package). Expectations of the service could be different if the choice was part of a package of services—e.g., part of the service experience could be attributed to the package.

2.1.2.6. Disconfirmation of search time in choosing the service (+). Disconfirmation theory suggests that overall satisfaction is likely to be lower when customers generally put in more effort than they expected to expend in choosing the service (Anderson & Sullivan, 1993; Oliver, 1980). If they actually do spend less time relative to their expectations, then we should see a higher level of overall satisfaction. However, if customers spend more time in choosing a service than they expected to (regardless of the medium), they will be less satisfied because their outcome was worse than expected.

In addition to the focal constructs and control variables that we have identified above, specific characteristics of a website may further modify the effect of the online environment on service encounter satisfaction. In particular, we consider two factors: (1) interactivity of website, and (2) depth of information at the website.

2.1.2.7. Interactivity of website (+). Interactivity refers to the ability of websites to dynamically generate outputs based on customer queries and searches. A static website is likely to increase search effort for online shoppers, thereby decreasing their satisfaction. On the other hand, a well designed interactive website could generate higher satisfaction by providing greater control to customers to personalize the information search. Greater customer control of the shopping experience increases the pleasure and convenience of shopping, an important component of customer satisfaction (Marmorstein, Grewal, & Fisher, 1992). These arguments suggest that the customer’s service encounter satisfaction is positively related to the degree of interactivity of the website from which the customers choose.

2.1.2.8. Depth of information at the website (+). Websites can offer more information at the point of choice, thereby helping customers make better choices. Some websites have multiple layers of web pages with detailed information. If the information is multi-layered and rich, customers will tend to value the service encounter more than when the information is very superficial (Glazer, 1991). Thus, a customer’s service encounter satisfaction is positively related to the depth of information available at the website (in our study, the service provider’s site) from which the service is chosen.

Among the control variables that influence service encounter satisfaction, only prior experience impacts overall satisfaction. The remaining variables are specific to a service encounter. Therefore, we do not include them as potential drivers of overall satisfaction. Finally, note that we predict a positive effect of service encounter satisfaction on overall satisfaction because overall satisfaction is driven by satisfaction.
from a series of service encounters (Bitner & Hubbert, 1994; Oliver, 1997; Rust & Oliver, 1994).

2.2. Drivers of loyalty

In Fig. 1, we have summarized the drivers of loyalty in a service industry, namely, the medium, overall customer satisfaction, ease of obtaining information, frequency of use, prior experience with service provider, membership in frequency program, and income. These factors are particularly applicable to the travel industry (Pritchard & Howard, 1997). We discuss the effects of each of these factors below.

2.2.1. Focal constructs

2.2.1.1. Direct (main) effect of the online medium (+). As suggested earlier, other things being equal, when customers make a service choice online, the resulting service is more likely to meet their expectations. Because the Web facilitates information search, we expect the customer to be better informed about the service levels to expect, and is less likely to be surprised (positively or negatively) by the services received when choosing online than offline. As a result, the customer’s confidence in the service provider increases, which builds “fortitude” that prevents encroachment by competitive brands (Oliver, 1999). Further, customers may bookmark the URL of service providers when they have increased confidence in a service provider (i.e., reduce the consideration set), which makes it easier to find the same provider during future purchase occasions. The online medium also allows the customer to inspect more fully the available choices (e.g., pictures of hotels) than the offline medium. This makes it easier for a customer to learn about available options faster online than offline, which could result in “cognitive lock-in” to the selected service provider by making it easier for the customer to deal with that service provider than with other providers (Johnson et al., 2002). Finally, the ability to choose service providers on a 24/7 basis lowers the probability that the customer would go to another service provider simply because of limited access to his or her preferred service provider.

On the other hand, one could also argue that the online medium facilitates easier comparison of alternatives and allows faster propagation of any negative word-of-mouth than the offline medium. These factors could increase the consideration sets of customers, thereby increasing the chances that customers gain more information/positive attitude toward competing services, and loss of loyalty to the focal service. However, since service failures typically constitute a small proportion of all service encounters, we expect the positive influences of these factors to dominate the negative influences. Indeed, a significant majority of customers prefer to reduce the number of items they consider online even when there are a large number of alternatives (Wu & Rangaswamy, in press). In sum, we expect the online medium to directly enhance customer loyalty, which we state in the form of the following hypothesis.

H5: Other things equal, loyalty to the service provider is higher when the service is chosen online than offline.

2.2.1.2. Overall customer satisfaction (+). Customers’ loyalty to a service provider is influenced by their overall satisfaction with that provider. Most prior research has found qualified support for a positive satisfaction–customer retention relationship (Rust & Zahorik, 1993). Bolton (1998) found that the duration of a relationship between a customer and a service provider is longer when the customer is satisfied. Crosby and Stephens (1987) found that prior satisfaction increases the likelihood of a customer renewing her insurance policy. Rust, Zeithaml, and Lemon (2000) propose that the relationship between satisfaction and loyalty is positive, i.e., the more satisfied customers are with a service provider the more loyal they are to the service provider. Reicheld (1996), however, suggests that satisfaction is significantly related to loyalty only at very high levels of satisfaction. Therefore, the exact nature of the relationship between overall satisfaction and loyalty is an empirical issue.

At the same time, we expect overall customer satisfaction to have stronger effect on loyalty online than offline. One reason for this differential effect is that when customers are satisfied with a service provider online, they could more readily bookmark the website, make the choice 24×7, and even store or place a future order with the service provider. Thus, they have to spend less effort online for making a choice decision
of roughly the same quality as their previous choices. Thus, the online medium intensifies the relationship between satisfaction and loyalty (Johnson & Payne, 1985). A further reason for the online medium to reinforce the relationship between satisfaction and loyalty is that a satisfied customer could more easily locate the same service even in an unfamiliar context (e.g., a favorite hotel chain in a new city). These arguments lead to the following hypothesis.

**H₆:** The positive effect of overall satisfaction on loyalty to a service provider is greater for customers who choose the service online than it is for those who choose offline.

2.2.1.3. **Ease of obtaining information (+).** Like overall customer satisfaction, loyalty may also depend on the ease with which relevant information can be obtained while choosing the service provider. Easier access to information typically reinforces the tendency to go back to a preferred service provider (Oliver, 1999). As we suggested earlier, for products such as hotels that have a large number of search attributes, it is easier to obtain information online than offline. Ease of obtaining information also increases customer learning and may lead to greater cognitive lock-in to the service online than offline (Johnson et al., 2002). Thus, we expect higher loyalty online due to this factor. This reasoning leads to H₇.

**H₇:** The positive effect of ease of obtaining information on loyalty for a service provider is higher for customers who choose the service online than it is for those who choose offline.

2.2.2. **Control variables**

2.2.2.1. **Prior experience (+).** A customer who has a favorable prior experience with a service provider will likely be loyal to that service provider (Reicheld, 1996). This effect is likely to be higher in the online environment because it is more convenient and faster to choose the same service online than offline (e.g., through the use of bookmarks).

2.2.2.2. **Frequency of service use (+).** Both positive and negative relationships between frequency of use and loyalty are possible (Rust et al., 2000). It is not clear either whether the effect of frequency of use on loyalty in the online environment will be any different from that in the offline environment.

2.2.2.3. **Membership in frequency program (+).** If a person is a member of a service provider’s frequency program, he/she typically benefits more by choosing the service provider than if he/she is not a member. This aspect is likely to be associated with greater loyalty to the service provider.

2.2.2.4. **Income (+).** An individual’s income level may influence his/her loyalty to a service provider. Customers with lower discretionary incomes would be willing to do more price-comparisons and be less loyal to a service provider than those with higher incomes. Also, higher-income individuals typically have more time constraints and are more likely to stick to the same service provider.

3. **Data**

3.1. **Data collection procedure**

To test our hypotheses, we collected two sets of data in the lodging sector of the travel services industry, which comprises about 35% of online consumer sales (Comscore Report, 2002). These data sets are: (1) a sample of customers who used both an online (i.e., a website) and an offline medium (e.g., a travel agent or hotel reservation desk) for making two different hotel reservations (Data set 1), and (2) separate samples of customers of a single hotel chain (Data set 2), one sample of customers who made reservations online and another sample which made reservations offline.

Our use of the two different data sets in the same study is similar to developing multiple models (e.g., Silk & Urban, 1978) and conducting multiple experiments in consumer behavior research to assess convergence in results and insights. Data set 1 is akin to a within-subjects design, whereas Data set 2 is akin to a between-subjects design in an experimental study. In Data set 1, we have the advantage of focusing on the effect of the online medium relative to the offline medium for the same customer, thereby eliminating the variance due to differences in the online and
offline populations. In Data set 2, we have the advantage of having an identical service provider for both the online and offline samples, which reduces variance due to services delivered. Although we used identical questionnaires in both data sets, the questions in Data set 1 were applicable to any hotel chosen by the respondent, whereas the questions in Data set 2 were specific to Marriott hotels. Further, in Data set 1, the respondent answered the set of questions twice, once with respect to his/her most recent online reservation and once with respect to his/her most recent offline reservation. All respondents were provided identical incentives to participate, in the form of 500 frequent reward points for a member hotel of HSMAI and for Marriott hotel in Data sets 1 and 2, respectively.

3.1.1. Data set 1

We collected primary data by surveying a random sample of hotel customers using a list provided by the Hospitality Sales and Marketing Association International. Customers in this sample were representative of customers who chose hotels in both the online and offline media. We received 144 usable responses from a sample of 1000 customers who were randomly selected from the list and mailed questionnaires (a response rate of 14.4%). This sample contained only those who had chosen a hotel both online and offline. The same respondents provided data for their most recent offline and online reservations.

The online and offline questionnaires included items on shopping behavior, choices, attitudes, measures of the hypothesized factors, and demographics. The questions relating to factors common to both the online and offline reservations were identical in the two questionnaires. In addition, the online questionnaire included items on website factors. Most of the questions related to the most recent service encounter. This way, we were able to elicit attitudes with respect to a specific experience. To ensure that the information search undertaken by the respondent is done within a single medium (online or offline), we included a screening question regarding the medium from which the respondent obtained information.

3.1.2. Data set 2

We obtained primary data from both online and offline customers of Marriott International. The online respondents chose a Marriott hotel on the Internet and the offline respondents chose a Marriott hotel using conventional offline methods (e.g., travel agents or toll-free telephone number). We screened the offline customers, and none of them had any prior online reservation experience. By having two separate samples, we are able to compare the attitudes and behaviors of populations that differ primarily in the medium they use to make hotel reservations.

The offline sample consisted of respondents from a random sample of 2000, generated from the population of 12 million customers of Marriott International’s customer database, who were mailed the questionnaire. For the online sample, we posted the survey at Marriott’s website with a “New” tag and requested that the online visitors fill out the survey.

Note that it is extremely difficult to achieve random sampling on the Internet because of the lack of a sampling frame—there is no list of all Internet users or even a list of potential visitors to a site, nor is there an online equivalent to random digit dialing. However, it is worth noting that online surveys typically elicit responses from a reasonably representative sample of the universe of online shoppers of a given product. This is because those who are likely to shop online for a product also tend to be the ones who respond to online surveys on the product (Intelliquest, 1997).

We collected 190 usable responses from the online survey and 403 usable responses from the offline survey, which represented a response rate of about 20.2%. As is well recognized, the online population is younger, on average, than the general population. Of the responses we obtained from the offline surveys, we selected the first 272 that allowed us to closely match the online sample in demographics. This sam-
ple formed the matched offline sample. To match the online and offline samples on demographics, therefore, we dropped several respondents in the older age groups from the original offline sample. Instead of using a simple random sample to identify respondents to be dropped, we chose to retain respondents in the order in which we received their completed surveys—keeping respondents who sent their surveys early. We did this because early respondents are more likely to have a “response tendency” similar to those visitors to a website who choose to fill out the online questionnaire (Intelliquest, 1997). Therefore, the matched offline sample allowed us to control for some observable individual differences and compare the attitudes and behaviors of populations that differed primarily with regard to the medium they used to make reservations.

6 Even so, we subsequently reanalyzed our data using a simple random sample to identify the respondents to be dropped, but the results of our empirical analysis were not much different.

7 We were not able to collect multiple-item measures for some of our constructs because of questionnaire length restrictions. There were several other questions (not directly relevant to our study) that were also included in the survey and were driven by the management needs of HSMAI and Marriott in Data sets 1 and 2, respectively. Lack of multiple-item measures is not necessarily disadvantageous in services research (Drolet and Morrison, 2001).

Table 2 summarizes the demographics of the online and offline samples in both data sets.

In Data set 1, about a quarter of the respondents were women, 58% were in the 35–54 age group, 85% had college education or above, and about 79% had annual family income of $60,000 or above. In Data set 2, comparing the online sample with the unmatched or naturally occurring offline sample, we see that the samples are very similar in terms of gender composition, proportion of 35–54 age group, and the percentage of college-educated members. However, on average, online customers are younger, less educated, and less affluent than the unmatched offline customers. Table 2 also summarizes the characteristics of the offline matched samples in Data set 2 that we used for all our data analyses. The online and matched offline samples are very comparable in terms of gender distribution, age, education level, and annual family income.

Tables 3a and 3b summarize the variables we use in our analysis, their operationalizations, and their mean values for the samples in Data sets 1 and 2, respectively. The tables also indicate the variables whose mean values are significantly different online and offline.

Table 3c summarizes the correlations among the key variables in both the data sets. Several key correlations (e.g., between satisfaction and loyalty)
are higher online than offline. The satisfaction—loyalty correlation ranges from 0.42 to 0.52 across the samples. Tables 3a–3c suggest remarkable similarities between the two data sets on most variables, even though the samples in these two data sets are
Table 3b
Operationalization, means and variances of variables in data set 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Operationalization</th>
<th>Online mean (Var)</th>
<th>Offline: unmatched mean (Var)</th>
<th>Offline: matched mean (Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service encounter satisfaction</td>
<td>Expressed satisfaction on the most recent service encounter with the service provider (5-point scale: Very dissatisfied – Very satisfied)</td>
<td>4.42 (0.69)</td>
<td>4.39 (0.58)</td>
<td>4.38 (0.62)</td>
</tr>
<tr>
<td>Overall satisfaction (OVRSAT)</td>
<td>Expressed level of overall satisfaction with the service provider (7-point scale)</td>
<td>5.78 (1.46)</td>
<td>5.70 (1.15)</td>
<td>5.72 (1.02)</td>
</tr>
<tr>
<td>Loyalty (LOYAL)</td>
<td>Expressed degree of loyalty to the service provider (7-point scale)</td>
<td>5.73 (2.23)</td>
<td>5.20*** (1.92)</td>
<td>5.26*** (1.77)</td>
</tr>
<tr>
<td>Ease of obtaining information (EASINFO)</td>
<td>Average rating on the ease of obtaining information on several factors relevant for the choice of hotel (e.g., location, physical condition, etc.) on 5-point scale: “Very difficult to Very easy”</td>
<td>3.75 (0.35)</td>
<td>3.79 (0.35)</td>
<td>3.79 (0.39)</td>
</tr>
<tr>
<td>Frequency of use (FREQUSE)</td>
<td>Expressed frequency of use for the product (scale ranging from “One stay every 3 months or more to “One stay a week”)</td>
<td>2.35 (1.90)</td>
<td>2.03** (1.53)</td>
<td>2.10* (1.53)</td>
</tr>
<tr>
<td>Rating on reputation (RATREP)</td>
<td>Assessment of reputation of the service provider (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.91 (0.73)</td>
<td>3.92 (0.59)</td>
<td>3.95 (0.51)</td>
</tr>
<tr>
<td>Rating on room amenities (RATROOM)</td>
<td>Assessment of room amenities on the last service encounter (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.58 (0.83)</td>
<td>3.58 (0.62)</td>
<td>3.58 (0.62)</td>
</tr>
<tr>
<td>Rating on customer service (RATSERV)</td>
<td>Assessment of customer service on the last service encounter (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.62 (0.89)</td>
<td>3.67 (0.73)</td>
<td>3.67 (0.80)</td>
</tr>
<tr>
<td>Rating on price (RATPRICE)</td>
<td>Assessment of price on the last service encounter (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.37 (0.73)</td>
<td>3.33 (0.71)</td>
<td>3.32 (0.65)</td>
</tr>
<tr>
<td>Rating on location (RATLOC)</td>
<td>Assessment of hotel location on the last service encounter (5-point scale: “Well below average” to “Well above average”)</td>
<td>4.11 (0.68)</td>
<td>4.15 (0.57)</td>
<td>4.14 (0.60)</td>
</tr>
<tr>
<td>Rating on ease of reservations (RATEASE)</td>
<td>Assessment of ease of making reservations on the last service encounter (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.89 (0.80)</td>
<td>3.69** (0.83)</td>
<td>3.65** (0.85)</td>
</tr>
<tr>
<td>Prior experience (PREXP)</td>
<td>Perception of past experience with the service provider (5-point scale: “Well below average” to “Well above average”)</td>
<td>3.71 (1.07)</td>
<td>3.68 (0.88)</td>
<td>3.67 (0.86)</td>
</tr>
<tr>
<td>Customer’s disconfirmation of time to choose (TIMEDIFF)</td>
<td>Time “willing to spend” choosing a hotel less actual time typically spent (5-point scales ranging from “5 minutes or less” to “More than 60 minutes”)</td>
<td>0.16 (1.29)</td>
<td>0.55*** (1.72)</td>
<td>0.51** (1.73)</td>
</tr>
<tr>
<td>Perceived interactivity of message (INTERACT)</td>
<td>Perception of interactivity of the website message (5-point scale: “Much worse than at other sites” to “Much better than at other sites”)</td>
<td>3.71 (0.80)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Perceived depth of information (DEPTH)</td>
<td>Perception of extent of website information relative to expectations (5-point scale: “Much worse than expected” to “Much better than expected”)</td>
<td>3.85 (0.78)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Significantly different from online at 0.05 level.
** Significantly different from online at 0.01 level.
*** Significantly different from online at 0.001 level.
completely different. There are, however, a few notable differences in means between Data sets 1 and 2. For example, the frequency of use data suggests that the sample in Data set 2 consists of less frequent travelers. Although such differences in means between the two data sets are potentially interesting, our focus is primarily in exploring the relationship between customer satisfaction and loyalty, i.e., we focus on explaining why and how loyalty is higher online than offline and how it is related to customer satisfaction.

4. Models

To test our hypotheses, we develop three different sets of models.

- A simultaneous equation model in which satisfaction and loyalty are interrelated, such that each can be the driver of the other. This model is separate for online and offline samples and provides the base model for the next model, which we use to test the hypotheses.
- To formally test the effect of the online medium and the different hypotheses, we develop pooled models for the simultaneous equation and the recursive-system of equations, in which the medium is used as an additional (dummy) variable. This model is central to our analysis.
- A recursive system of three equations in which service encounter satisfaction influences overall satisfaction, which, in turn, influences loyalty. This model is separate for online and offline samples and forms the alternative model.

We describe each of these models below.

4.1. Simultaneous equation model of satisfaction and loyalty

According to our hypotheses, service encounter satisfaction, overall customer satisfaction and loyalty may be interrelated in that they drive one another. Such complex relationships can be captured by a simultaneous system of equations. We define the model for service encounter satisfaction as follows.\(^8\)

\[
\text{SESAT}_i = \beta_0 + \beta_1 \text{LOYAL}_i + \beta_2 \text{RATREP}_i + \beta_3 \text{RATROOM}_i + \beta_4 \text{RATSERV}_i + \beta_5 \text{RATPRICE}_i + \beta_6 \text{RATLOC}_i + \beta_7 \text{RATEASE}_i + \beta_8 \text{FREQUSE}_i + \beta_9 \text{PREXP}_i + \beta_{10} \text{CONST}_i + \beta_{11} \text{INTM}_i + \beta_{12} \text{PACK}_i + \beta_{13} D_i \cdot \text{INTERACT}_i + \beta_{14} D_i \cdot \text{DEPTH}_i + \epsilon_i
\]

where \(i\) is customer, SESAT is service encounter satisfaction, LOYAL is the loyalty to the service provider, RATREP is the customer rating on hotel reputation, RATROOM is the rating on room amenities, RATSERV is the rating on customer service, RATPRICE is the rating on price paid, RATLOC is the rating on location, RATEASE is the rating on ease of making a reservation, FREQUSE is the frequency of staying at hotels, PREXP is the rating on prior experience with the hotel, CONST is a dummy variable denoting whether traveler \(i\)'s hotel choice was driven by organizational constraints (1 if the choice was constrained, 0 otherwise), INTM is a dummy variable denoting if the traveler went through a travel website or a travel agent or intermediary before coming to the site of the hotel (1 if travel intermediary was involved, 0 otherwise), PACK is a dummy variable.

\(^8\) One factor often used in understanding price sensitivity in the travel industry is the purpose of travel, business, or leisure. However, there are no strong theoretical reasons why purpose of travel might influence customer satisfaction or loyalty. Even so, a subsequent empirical analysis that included the purpose of travel did not show it to be a significant factor influencing customer satisfaction or loyalty.

---

<p>| Table 3c Correlations among key variables in the data sets |
|---------------------------------|---------------------------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Online</td>
<td>Offline</td>
</tr>
<tr>
<td></td>
<td>Online</td>
<td>Offline</td>
</tr>
<tr>
<td>SESAT and OVSAT</td>
<td>0.61</td>
<td>0.40</td>
</tr>
<tr>
<td>SESAT and LOYAL</td>
<td>0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>OVSAT and LOYAL</td>
<td>0.52</td>
<td>0.46</td>
</tr>
<tr>
<td>LOYAL and FREQUSE</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.59</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>0.17</td>
</tr>
</tbody>
</table>
indicating whether the hotel choice was part of a travel package (1 if the choice was part of a travel package, 0 otherwise), INTERACT is the perceived interactivity of the website, DEPTH is the perceived depth of information on the website, \( D \) is a dummy variable =1 if the sample is online and =0 if the sample is offline, and \( \varepsilon \) is an error term assumed to be normally distributed with a mean of zero. This model captures the fact that satisfaction in a particular encounter may be driven by loyalty to the service provider.

Next, we propose the following model of overall satisfaction.

\[
OVRSAT_i = \beta_0 + \beta_1 SESAT_i + \beta_2 LOYAL_i + \beta_3 EASINFO_i + \beta_4 FREQUSE_i + \beta_5 PREXP_i + \beta_6 TIMEDIFF_i + \xi_i
\]

where \( OVRSAT \) is the overall satisfaction with the service provider, EASINFO is the general or overall ease of obtaining information on the alternatives, TIMEDIFF is the disconfirmation of search time in choosing the service (i.e., difference between the time generally spent in choosing a hotel and the time willing to spend in choosing a hotel), \( \xi \) is an error term assumed to be normally distributed with a mean of zero, and the other terms are as defined earlier.\(^9\) In this model, we allow overall satisfaction to influence loyalty.\(^10\)

4.2. Pooled simultaneous equation models customer satisfaction and loyalty

To formally test the effect of the online medium on the relationships, we propose pooled models of service encounter satisfaction, overall satisfaction, and loyalty. We make the parameters \( a_k, k \in \{0, 1, \ldots, 12\} \) in Eq. (1), \( b_l, l \in \{0, 1, \ldots, 6\} \) in Eq. (2), and \( \lambda_m, m \in \{0, 1, \ldots, 6\} \) in Eq. (3) a function of the medium (online or offline) as follows. For \( \phi_k \in \{\beta_k, \lambda_k\} \)

\[
\phi_k = \phi_k^F + \phi_k^N D_i
\]

where \( D_i = 1, \) when the data is for the online sample, and 0 otherwise.

Substituting Eq. (4) into each of Eqs. (1)–(3) produces the system of pooled models for our analysis.

4.3. Alternative model: recursive model of customer satisfaction and loyalty

To examine the incremental value of analyzing the relationship between customer satisfaction and loyalty as a system, we compared the results from the simultaneous equation system with those of an alternative model, i.e., a recursive model in which the \( LOYAL \) variable is absent in the equations for \( SESAT \) and \( OVRSAT \). When these variables are absent, the model becomes recursive and the three equations in such a model can be estimated independently by OLS (Dhrymes, 1974).

\(^9\) We do not include CONST, the variable indicating whether the hotel was chosen under organizational constraints, because it is a service encounter-specific variable and our dependent variable is overall loyalty. Even so, to test if this variable significantly drives loyalty, we estimated another model by including it. Its effect was insignificant, so we do not include it in our analysis.

\(^{10}\) To test for a possible nonlinear effect of overall satisfaction on loyalty as suggested by Reicheld (1996), we tried to use an additional quadratic term in overall satisfaction. This term, however, is highly correlated with overall satisfaction in our data, precluding the testing of a possible nonlinear effect in our analysis.
4.4. Model estimation

We estimate the set of pooled model equations using the three-stage least squares (3SLS) method. We also estimate Eqs. (1)–(3) as a simultaneous system of equations for each sample, namely, the online sample and offline sample in Data set 1 and the online sample, unmatched offline sample, and matched offline sample in Data set 2.

We tested all models for heteroscedasticity using the Glesjer (1969) test. The tests did not indicate the presence of heteroscedasticity in any of the models. Also, because the same respondent chose both online and offline in Data set 1, there is a possibility that his/her responses are correlated across the media. To account for this possibility, we allowed for errors from the online and offline samples to be correlated in the estimation. Before pooling, we tested for homogeneity of slopes and intercepts using the Chow (1960) test. The null hypothesis of homogeneity of parameters was rejected (\( p<0.001 \)).

5. Results and discussion

The results are summarized in Tables 4–6. The system of equations has good fit for cross-sectional data as indicated by the system-wide \( R^2 \) of 0.39 and 0.50 in the two data sets. The cross-residual correlations across the three equations are reasonably high (ranging from 0.43 to 0.73), indicating the appropriateness of the use of a simultaneous equation model for both data sets. The relative importance of a variable is indicated by its beta weight in each model. The results of the simultaneous system of equations separately for the online and offline samples for both data sets are consistent, overall, with those of the pooled simultaneous system of equations.11

5.1. Drivers of service encounter satisfaction and overall satisfaction

In both the data sets, service encounter satisfaction is not statistically different online versus offline. In Data set 1, however, overall satisfaction is higher when a hotel is chosen online than when selected offline (\( p<0.05 \)).

5.1.1. Direct effect of the online medium

After controlling for the effects of the service attributes, prior experience, frequency of service use, organizational constraints, and package choice, the online medium has no significant main effect on service encounter satisfaction in either data set. The online medium has no significant main effect on

| Table 4 |
| Results of the pooled model of service encounter satisfaction for online and offline customers |
| Variable (parameter) | Data set 1 | Data set 2 |
| | Pooled model | Beta | Pooled model | Beta |
| | (S.E.) | weight | matched | (S.E.) | weight |
| Intercept \( (z_0) \) | 1.34 | (0.42)*** | 1.50 | (0.61)*** |
| Focal variables |
| Online medium \( (z_0) \) | NS | NS |
| Loyalty—incremental online \( (z_1) \) | NS | NS |
| Control and other variables |
| Reputation \( (z_2) \) | 0.09 | (0.03)** | 0.12 | (0.08)* |
| Room amenities \( (z_3) \) | 0.21 | (0.07)*** | 0.20 | (0.07)*** |
| Service \( (z_4) \) | 0.23 | (0.12)* | 0.25 | (0.06)*** |
| Location \( (z_6) \) | 0.03 | (0.01)** | 0.02 | NS |
| Depth of information at the website \( (z_{14}) \) | 0.15 | (0.05)*** | 0.26 | (0.06)** |
| Sample size | 144 | 462 |
| System wide \( R^2 \) | 0.39 | 0.50 |
| System wide RMSE | 0.53 | 1.32 |

NS—not significant.
For expositional clarity, the control and other variables associated with nonsignificant results and the beta weights of nonsignificant variables are not shown.

\* \( p \leq 0.10 \).
\** \( p \leq 0.05 \).
\*** \( p \leq 0.01 \).

11 These results are not presented due to space limitations. They can be obtained by writing to the first author.
overall satisfaction either \((p<0.10)\). Thus, hypotheses \(H_{1a}\) and \(H_{1b}\) are not supported. Although the results do not support our hypotheses, they are consistent with what we should expect given the data—the mean levels of service encounter satisfaction are not different across the online and offline samples. The result regarding overall satisfaction is not surprising either in Data set 2 given that the mean values of service encounter satisfaction in the various samples are not statistically different. With respect to Data set 1, although the mean value of overall customer satisfaction in the online sample is significantly higher than it is in the offline sample, the analysis shows that this difference is not due per se to the medium. Rather it is primarily due to the other factors we included in the model. Note that this data involves multiple service providers and websites.

### Table 5
Results of the pooled model of overall customer satisfaction for online and offline customers

<table>
<thead>
<tr>
<th>Variable (parameter)</th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled model: (S.E.)</td>
<td>Beta weight (S.E.)</td>
</tr>
<tr>
<td><strong>Intercept ((b_0^F))</strong></td>
<td>(-0.90) ((0.53)^*)</td>
<td>(0.09) ((0.62))</td>
</tr>
<tr>
<td><strong>Focal variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online medium ((b_0^N))</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Loyalty—incremental online ((b_1^N))</td>
<td>(0.31) ((0.13)^**)</td>
<td>(0.26) ((0.13)^**)</td>
</tr>
<tr>
<td>Ease of obtaining information—incremental online ((b_2^N))</td>
<td>(0.05) ((0.02)^***)</td>
<td>NS</td>
</tr>
<tr>
<td>Frequency of use—incremental online ((b_3^N))</td>
<td>(0.15) ((0.08)^*)</td>
<td>(0.13) ((0.07)^*)</td>
</tr>
<tr>
<td><strong>Control and other variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service encounter satisfaction ((b_4^F))</td>
<td>(0.50) ((0.21)^**)</td>
<td>(0.64) ((0.13)^***)</td>
</tr>
<tr>
<td>Loyalty ((b_2^F))</td>
<td>(0.61) ((0.33)^*)</td>
<td>(0.73) ((0.17)^***)</td>
</tr>
<tr>
<td>Frequency of use ((b_3^F))</td>
<td>(-0.25) ((0.09)^**)</td>
<td>(-0.21) ((0.07)^***)</td>
</tr>
<tr>
<td>Prior experience ((b_4^F))</td>
<td>NS</td>
<td>(-0.11) ((0.07)^*)</td>
</tr>
<tr>
<td>Prior experience—incremental online ((b_5^F))</td>
<td>NS</td>
<td>(0.17) ((0.09)^*)</td>
</tr>
<tr>
<td>Sample size</td>
<td>144</td>
<td>462</td>
</tr>
<tr>
<td>System wide</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td>System wide (R^2)</td>
<td>0.53</td>
<td>1.32</td>
</tr>
</tbody>
</table>

NS—not significant.

For expositional clarity, the control and other variables associated with nonsignificant results and the beta weights of nonsignificant variables are not shown.

\(* p \leq 0.10.\)

\(** p \leq 0.05.\)

\(*** p \leq 0.01.\)

### Table 6
Results of the pooled model of loyalty for online and offline customers

<table>
<thead>
<tr>
<th>Variable (parameter)</th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled model: (S.E.)</td>
<td>Beta weight (S.E.)</td>
</tr>
<tr>
<td><strong>Intercept ((b_0^F))</strong></td>
<td>(1.07) ((0.36)^***)</td>
<td>(0.79) ((0.93))</td>
</tr>
<tr>
<td><strong>Focal variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online medium ((b_0^N))</td>
<td>(0.43) ((0.17)^**)</td>
<td>(0.41) ((0.22)^*)</td>
</tr>
<tr>
<td>Overall satisfaction—incremental online ((b_1^N))</td>
<td>(0.32) ((0.09)^**)</td>
<td>(0.50) ((0.20)^**)</td>
</tr>
<tr>
<td>Ease of obtaining information—incremental online ((b_2^N))</td>
<td>(0.10) ((0.04)^*)</td>
<td>(0.54) ((0.20)^**)</td>
</tr>
<tr>
<td><strong>Control and other variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall satisfaction—offline ((b_1^F))</td>
<td>(0.33) ((0.13)^***)</td>
<td>(0.43) ((0.19)^***)</td>
</tr>
<tr>
<td>Ease of obtaining information—offline ((b_2^F))</td>
<td>NS</td>
<td>(0.28) ((0.14)^*)</td>
</tr>
<tr>
<td>Frequency of use—offline ((b_3^F))</td>
<td>(0.24) ((0.13)^*)</td>
<td>(0.33) ((0.06)^***)</td>
</tr>
<tr>
<td>Prior experience—offline ((b_4^F))</td>
<td>(0.09) ((0.05)^**)</td>
<td>NS</td>
</tr>
<tr>
<td>Sample size</td>
<td>144</td>
<td>462</td>
</tr>
<tr>
<td>System wide (R^2)</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td>System wide (R^2)</td>
<td>0.53</td>
<td>1.32</td>
</tr>
</tbody>
</table>

NS—not significant.

For expositional clarity, the control and other variables associated with nonsignificant results and the beta weights of nonsignificant variables are not shown.

\(* p \leq 0.10.\)

\(** p \leq 0.05.\)

\(*** p \leq 0.01.\)
which may introduce idiosyncratic concerns regarding security, privacy, technology failure, and the like. These effects may have offset the favorable effects of the online medium on satisfaction, leading to a finding of no differences between online and offline customer satisfaction. As we indicated earlier, the actual main effects of the online medium can only be determined empirically.

5.1.2. Loyalty

Loyalty is significantly higher online than offline \((p<0.05)\) in both data sets. Loyalty has no significant effect on service encounter satisfaction in both the data sets \((p>0.10)\), therefore, \(H_{2a}\) is not supported. A specific service encounter may be treated as transactional, not relationship-oriented, so that both loyal and nonloyal customers may have similar reactions to similar service levels during a single service encounter. However, the positive impact of loyalty on overall customer satisfaction is incrementally higher in the online sample in both Data sets 1 and 2 \((\beta_{L}^{N}=0.31, p<0.05 \text{ in Data set 1}; \beta_{L}^{N}=0.26, p<0.05 \text{ in Data set 2})\), supporting \(H_{2b}\). This is an important result from our study. It suggests that attitudinal loyalty to the service provider is also a major factor in creating enduring satisfaction across service encounters, particularly in the online environment.\(^{12}\) Our result is consistent with Oliver (1999) who suggests that loyalty fortifies satisfaction toward the service provider, and other studies that have found loyal customers to have a lower decay rate of overall satisfaction, and to be more forgiving of less satisfactory services (Rust et al., 1999; Zeithaml, Berry, & Parasuraman, 1993). Our results also help explain the findings by Bain and Company which shows that the returns to loyalty-building initiatives are in the double-digits (Baveja, Rastogi, Zook, Hancock, & Chu, 2000), and lend further support to the results reported by Rust et al. (2000) who show that under certain conditions, it may be better to focus directly on loyalty-building initiatives (e.g., frequent-user programs), rather than work on satisfaction-enhancing efforts (e.g., total quality management).

5.1.3. Ease of obtaining information

The ease of obtaining information is higher online than offline in Data set 1 \((p<0.001)\), as expected, but there is no difference in Data set 2. The effects of ease of obtaining information are consistent—the easier it is to obtain information the higher the overall satisfaction, consistent with the results of Oliva et al. (1992). Ease of obtaining information has a greater positive impact on overall customer satisfaction when the service is chosen online than offline \((\beta_{0}^{N}=0.05, p<0.01 \text{ in Data set 1})\), supporting \(H_3\). This is consistent with Johnson and Payne (1985) in that the online medium has lower cognitive costs than the offline medium, leading to greater customer utility and satisfaction.

5.1.4. Frequency of use

Frequency of use is not significantly different between online and offline shoppers in Data set 1 but is higher online than offline in Data set 2 \((p<0.05)\). Frequency of use is negatively related to overall satisfaction in both data sets. These results support prior findings (Bolton & Lemon, 1999; Vredenburg & Wee, 1986). More frequent users have higher expectations. Thus, they are less likely to be satisfied than infrequent users.\(^{13}\) The online medium, however, seems to mitigate this negative relationship—its incremental effects are positive in both data sets \((\beta_{4}^{N}=0.15, p<0.05 \text{ in Data set 1}; \beta_{4}^{N}=0.13, p<0.05 \text{ in Data set 2})\), consistent with \(H_4\).

5.2. Drivers of loyalty

5.2.1. Direct effect of the online medium

Loyalty is significantly higher online than offline \((p<0.05)\) in both data sets. A key result of our analysis is that the online medium has a positive impact on loyalty of the types of people in our two data sets, even after accounting for other factors that influence loyalty \((\hat{\beta}_{0}^{N}=0.43; p<0.05 \text{ in Data set 1}; \hat{\beta}_{0}^{N}=0.41; p<0.10 \text{ in Data set 2})\), supporting \(H_5\). In Data set 2, in particular, there are almost no differences between the online and offline samples on product attributes (this

\(^{12}\) Although our data do not allow us to make precise determination of the direction of causality, the simultaneous equation approach reduces the possibility that this is simply a correlation between loyalty and overall satisfaction.

\(^{13}\) Respondents in Data set 1 had higher average frequency of use and lower satisfaction levels (on service encounter satisfaction and overall satisfaction) than respondents in Data set 2 (Tables 3a and 3b).
is expected given that both samples booked the same hotel, namely, Marriott) and yet, there is an incremental effect of the online medium on loyalty. Perhaps the online medium increases loyalty by making it easier for satisfied customers to choose the same hotel again (e.g., through bookmarks).

5.2.2. Overall satisfaction

Overall satisfaction enhances loyalty in both the samples in Data sets 1 and 2 ($\lambda_1^1 = 0.33, p < 0.05$ in Data set 1; $\lambda_1^2 = 0.43, p < 0.01$ in Data set 2). This finding supports similar findings (e.g., Bolton, 1998; Rust & Zahorik, 1993) regarding the effects of satisfaction on behavioral loyalty (or customer retention) and extends those results to attitudinal loyalty. More interestingly, the positive relationship between satisfaction and loyalty is stronger online than offline in both data sets ($\lambda_1^N = 0.32, p < 0.05$ in Data set 1; $\lambda_1^N = 0.50, p < 0.05$ in Data set 2), thereby supporting hypothesis H6. Therefore, it appears that satisfied customers are better able to express their loyalty online than offline, perhaps through the use of bookmarks, search features, and hot links associated with the content presented at the website.

5.2.3. Ease of obtaining information

Ease of obtaining information has a greater positive impact on loyalty when the service is chosen online than offline ($\beta_2^N = 0.10, p < 0.05$ in Data set 1; $\beta_2^N = 0.54, p < 0.01$ in Data set 2), supporting H7. This is consistent with Oliver (1999) who suggests that ease of information search and retrieval generally increases the customer’s likelihood of returning to the same service. The greater impact online relative to offline may be due to the learning that occurs when information search is easier, resulting in greater cognitive lock-in to the service provider in the online environment (Johnson et al., 2002). This result, together with the condition that search costs are typically lower online than offline, suggests that website content and the ability to access them at the time of decision-making favorably impact loyalty online.

5.3. Effects of other variables, including control variables

Among the control variables, customer perceptions of the attributes of the service encounter such as reputation, room amenities, customer service, and ease of reservations are not statistically different online versus offline. Customers perceived that they got a more favorable price online than offline (Data set 1), their location choices were better offline than online (Data set 1), that it was easier to make reservations online than offline (Data set 2), and that those who had made the hotel choices online had better prior experience with the service provider (Data set 1) and were willing to spend more time online than offline when choosing hotels (both data sets). The actual time spent choosing a hotel online was significantly higher at 2.97 versus 2.19 offline ($p < 0.001$) on a 6-point scale, where 1 was anchored at less than 5 min and 6 was anchored at more than 60 min (this data is not included in Tables 3a–3c because we do not use this variable in our models). This result was also replicated in Data set 2 (2.71 online versus 1.84 offline; $p < 0.01$). Depth of information at the website has a positive impact on service encounter satisfaction ($p < 0.05$), but interactivity of the website is not significant in either data set. Apparently, content is more important than interactivity in influencing customer satisfaction when customers make choices online, a finding consistent with what several Web gurus proclaim: “Ultimately, users visit your website for its content. Everything else is just the backdrop. The design is there to allow people access to the content.” (Nielsen, 2000). Also, at the time of this study, major hotel websites did not have rich interactivity features, such as personalization. Perhaps with increasing personalization features, website interactivity could have a stronger influence on service encounter satisfaction.

As expected, the service attributes have significant effects in the expected direction on service encounter satisfaction (Table 4). The fact that location has little or no influence on service encounter satisfaction in Data set 2 suggests that even if location is the most important attribute influencing a customer’s choice of a hotel, it does not determine satisfaction with the service provider. Several control variables (e.g., prior experience) have a statistically significant effect on overall satisfaction and loyalty, consistent with Bolton and Lemon (1999), Oliva et al. (1992), and Woodruff et al. (1983). Generally, these variables have a slightly larger impact on overall satisfaction and loyalty in Data set 2 than in Data set 1.
In Data set 2, because most online respondents (89%) and all offline respondents were frequent program members, we dropped this variable in the final model for overall satisfaction. To ensure consistency, we also dropped this variable for Data set 1, but this variable was not significant at the 0.01 level in the loyalty equation for Data set 1 even when we included it in the model.

As expected, service encounter satisfaction significantly enhances overall satisfaction both online and offline in both data sets ($\beta_1^O=0.50, p<0.05$ in Data set 1; $\beta_1^C=0.64, p<0.01$ in Data set 2). This finding is consistent with the belief that overall satisfaction is driven by satisfaction with a series of service encounters (Bitner & Hubbert, 1994; Oliver, 1997; Rust & Oliver, 1994).

5.4. Alternative model

To assess the incremental value of the simultaneous equation models, we compared the results from these models with those from the recursive models.14 The comparison highlighted the appropriateness and importance of considering loyalty as a driver of customer satisfaction. The model fits and the interpretability of parameter signs were better for the simultaneous equation model than for the recursive model ($R^2$ ranging from 0.28 to 0.42 and five significant parameters with the wrong signs for the recursive model). Further, loyalty was a significant driver of overall customer satisfaction in the simultaneous equation model ($p<0.05$), whereas loyalty had no effect on satisfaction in the recursive model. Recall that unlike the recursive model, the simultaneous system includes the impact of loyalty on both service encounter satisfaction and overall satisfaction, thus allowing us to explore the impact of loyalty on satisfaction.

5.5. Summary of results and study contributions

Table 7 summarizes the actual parameter signs obtained from our analyses along with the hypothesized parameter signs. Contrary to our expectations ($H_{1a}$ and $H_{1b}$), the online medium does not have a significant effect on either service encounter satisfaction or overall satisfaction. The positive effect of loyalty on overall satisfaction is higher online than offline ($H_{2b}$). Ease of obtaining information has a stronger positive effect on overall satisfaction online than offline ($H_3$). Frequency of use has a lesser negative effect on overall satisfaction online than offline ($H_4$). Loyalty to the service provider is inherently higher if customers choose services online ($H_5$). We think that the most interesting results from our study are that loyalty has a stronger positive reciprocal impact on overall satisfaction when customers make choices online and that overall satisfaction had stronger positive impact on loyalty online than offline ($H_6$). Finally, ease of obtaining information has a stronger positive effect on loyalty online than offline ($H_7$).

---

14 The results of the recursive models are not presented to conserve space. They are available from the first author.

---

Table 7: Hypotheses results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Actual sign</th>
<th>Predicted sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online medium on service encounter satisfaction ($H_{1a}$)</td>
<td>$\beta_1^O$</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Online medium on overall customer satisfaction ($H_{1b}$)</td>
<td>$\beta_1^C$</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Loyalty on service encounter satisfaction—incremental online ($H_{2a}$)</td>
<td>$\beta_2^O$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Loyalty on overall satisfaction—incremental online ($H_{2b}$)</td>
<td>$\beta_2^C$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ease of obtaining information on customer satisfaction—incremental online ($H_3$)</td>
<td>$\beta_3^O$</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Frequency of use on customer satisfaction—incremental online ($H_4$)</td>
<td>$\beta_4^O$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Online medium on loyalty ($H_5$)</td>
<td>$\beta_5^O$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Overall satisfaction on loyalty—incremental online ($H_6$)</td>
<td>$\beta_6^O$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ease of obtaining information on loyalty—incremental online ($H_7$)</td>
<td>$\beta_7^O$</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

NS—not significant.
Taken together, these results offer the following answers to the questions that motivated our research. We had hypothesized that the shopping medium (online or offline) should affect both the efficiency (effort, cost) and effectiveness (the actual hotel chosen) of shopping, and therefore, we should expect satisfaction with a service to be different when chosen online versus offline. Surprisingly, our results did not support this view, at least for the types of services that we studied. This suggests that people care more about the actual service received, which apparently is no different whether the service is chosen online or offline, i.e., customers make satisfaction-equivalent choices, on average, regardless of whether they chose the service provider (hotel) online or offline.

At the same time, customers exhibit greater loyalty when they choose a hotel online, and the relationship between overall satisfaction and loyalty is also higher when the service is chosen online. This suggests that either loyal customers shop online, or customers are better able to express their loyalty online. The first explanation is unlikely given that we find loyalty to be higher online even in Data set 1, and also because loyalty is not highly correlated with frequency of stay. There is also a stronger reinforcing relationship between loyalty and overall satisfaction in the online environment. Together, overall satisfaction has a stronger effect on loyalty online, and in turn, loyalty has a stronger effect on overall satisfaction. Finally, the online medium offers more information and makes it easy for customers to access and use that information, which also fosters loyalty to the service provider.

Our results have extended prior research in two important ways. First, to our knowledge, this is the first study to compare satisfaction and loyalty across online and offline environments. Although there are substantial similarities online and offline among the determinants of satisfaction and loyalty and their relationship (e.g., the effects of service attributes, the determinants of service encounter satisfaction), the study has identified some incremental effects (e.g., effects of the ease of obtaining information) and some unique effects (e.g., depth of information at a website) of the online medium. Second, our study has uncovered a more complex relationship between satisfaction and loyalty—these two variables have a reciprocal relationship between them, an issue that has not been explored in previous research.

6. Managerial implications, limitations, and extensions

6.1. Managerial implications

Based on our study, we recommend the following strategies and tactics for service providers.

6.1.1. Use the online medium to reinforce loyalty

Satisfaction builds loyalty, which reinforces satisfaction, a phenomenon which is stronger online than offline. Service marketers should consider developing special loyalty-enhancing initiatives for their online customers to reinforce overall satisfaction. For example, a hotel offering rewards based on number of hotel stays could (1) provide additional reward points for booking online, (2) prominently feature these rewards at its website, (3) enable customers to keep track of their reward positions, and (4) proactively remind or encourage customers to act when they get close to their reward milestones. While some companies have implemented the first three initiatives, not many are doing the fourth initiative.

6.1.2. Enhance the information content of the website

Our results show that the depth of information at the website increases service encounter satisfaction, which increases overall satisfaction, which, in turn, has a mutually reinforcing relationship with loyalty. Hotel marketers can enhance the depth of information at their websites by including such content as local weather, nearby attractions and restaurants, maps, events that will be taking place in a locality in the near future, a detailed list of amenities, pictures of rooms, etc. Yet another promising way to enhance content is for the service provider to partner with other service providers in related markets (e.g., a nearby car rental agency or nearby restaurants).

Although we have tried to get equivalent groups of online and offline customers in Data set 2, the population of online users could be systematically different from the population of offline consumers, so our results apply mainly to the sampling frames used in the study. Our results do not imply, for example, that service providers seeking to improve customer loyalty would necessarily benefit from a forceful conversion of offline customers to the online medium. In fact, forceful conversion of those customers who prefer to transact offline may likely diminish their loyalty.
6.1.3. Make information access as easy as possible on the website

The finding that ease of obtaining information has a stronger effect on both overall satisfaction and loyalty online than offline has important implications. First, a service provider should determine the appropriate information needs of its customers. Second, it should design a website so that the customer can access not only all the relevant information but also access them only when he/she needs it. Third, the company should invest in technologies that search for the right information and retrieve the information as quickly as possible. Fourth, the company may want to focus on creating the right user-interfaces that allow customers to access information in the way that is most convenient for them (e.g., on wireless device, without visuals, etc.).

6.1.4. Provide greater value to frequent online users

More frequent users seem to have greater overall satisfaction when they choose the service provider online than offline. This finding implies that companies should focus on increasing the frequency of use by online customers. These could be provided in the form of online promotions and incentives that offer more hotel stays or free hotel stays for increased use of the service.

6.2. Limitations and extensions

Our research has some limitations that should be addressed by future research. First, a major limitation is that we could not use multiple-item measures for most of our constructs due to data collection restrictions imposed by the study sponsors. It might be useful to replicate our study, perhaps in other service industries, using multi-item measures. Second, to enhance the generalizability of the findings, the study should be extended to other industries with a different set of contextual and competitive characteristics. For example, it would be interesting to study industries in which the entire service encounter including delivery is either online or offline (e.g., financial services). Third, it would be useful to explore whether loyalty to service provider extends to brand loyalty, an issue that would be particularly relevant for service providers with multiple brands. Fourth, an analysis of the impact of the medium on switching behavior would be a worthwhile complement to the analysis of loyalty and customer satisfaction. This analysis is typically done using longitudinal data. Finally, future research should explicate the customer decision processes by which alternative forms of loyalty, especially ultimate loyalty (Oliver, 1999), are formed and maintained both online and offline.

Acknowledgements

We thank Marriott International and HSMAI Foundation for providing the data. We also thank Hubert Gatignon, Ruth Bolton, Roland Rust, Richard Oliver, three anonymous reviewers, the participants at the 1999 INFORMS Marketing Science conference in Syracuse, New York, and the participants at the 2000 Marketing Camp at the University of Texas at Austin for their valuable comments.

References


Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction

Wagner A. Kamakura\textsuperscript{a,*}, Michel Wedel\textsuperscript{b,c}, Fernando de Rosa\textsuperscript{d}, Jose Afonso Mazzon\textsuperscript{e}

\textsuperscript{a}Fuqua School of Business, Duke University, Durham, NC 27708, USA
\textsuperscript{b}Faculty of Economics, University of Groningen, 9700 AV Groningen, Netherlands
\textsuperscript{c}University of Michigan Business School, 701 Tappan Street, Ann Arbor, MI 48109, USA
\textsuperscript{d}Universidade de Brasilia, SQSW 394 Bloco 1, Apto 507, Brasilia 70673-409, DF, Brazil
\textsuperscript{e}Universidade de S\~{a}o Paulo, Faculdade de Economia, Administra\c{c}\~{a}o e Contabilidade, 05508-900, S\~{a}o Paulo, Brazil

Received 1 August 2001; received in revised form 1 May 2002; accepted 14 May 2002

Abstract

An important aspect of the new orientation on customer relationship marketing is the use of customer transaction databases for the cross-selling of new services and products. In this study, we propose a mixed data factor analyzer that combines information from a survey with data from the customer database on service usage and transaction volume, to make probabilistic predictions of ownership of services with the service provider and with competitors. This data-augmentation tool is more flexible in dealing with the type of data that are usually present in transaction databases. We test the proposed model using survey and transaction data from a large commercial bank. We assume four different types of distributions for the data: Bernoulli for binary service usage items, rank-order binomial for satisfaction rankings, Poisson for service usage frequency, and normal for transaction volumes. We estimate the model using simulated likelihood (SML). The graphical representation of the weights produced by the model provides managers with the opportunity to quickly identify cross-selling opportunities. We exemplify this and show the predictive validity of the model on a hold-out sample of customers, where survey data on service usage with competitors is lacking. We use Gini concentration coefficients to summarize power curves of prediction, which reveals that our model outperforms a competing latent trait model on the majority of service predictions.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Database marketing; Cross-selling; Customer relationship management

1. Introduction

As many product and service markets become saturated and highly competitive, vendors realize that the acquisition of new customers happens mostly at the expense of competitors and, at the margin, these new customers tend to be “switchers” who will likely switch again in response to an attractive competitive offer. This competition for new customers in mature markets leads to the phenomenon known as “churn,” in which each vendor becomes a revolving door of acquired and lost customers. In order to escape this vicious circle, firms are increasingly focusing on...
strengthening the relationships with their customers (Day, 2000). Customer relationship management (CRM) has been more than a “buzzword” in management and marketing circles. According to industry sources, worldwide CRM-related investments reached $3.3 billion in 1999 and are expected to reach $10.2 billion by 2003.

One of the main CRM tools for forging stronger relationships with customers is cross-selling (Kamakura, Ramaswami, & Srivastava, 1991). The rationale for cross-selling as a strategy for reducing customer “churn” is very simple. As a customer acquires additional services or products from a vendor, the number of points where customer and vendor connect increases, leading to a higher switching cost to the customer. For example, it is easier for a customer with only a checking account to close this account than for another customer who also has automatic paycheck deposit and bill payments. Another important benefit of cross-selling, not as immediately visible as the increase in customer switching costs, is that it allows the firm to learn more about the customer’s preferences and buying behavior, thereby increasing its ability to satisfy the customer’s needs more effectively than competitors. For example, as a bank increases its “share-of-wallet” from a customer, it becomes more familiar with the customer’s financial needs, and in a better position than competitors to develop and offer services that satisfy those needs.

On the other hand, cross-selling can also potentially weaken the firm’s relationship with the customer, because frequent attempts to cross-sell can render the customer non-responsive or even motivated to switch to a competitor. In order to effectively cross-sell its products/services, the marketer must find—in commonly used jargon—the right offer for the right customer at the right time. The customer transaction database is instrumental in achieving that, because it allows the firm to learn about a customer, through its experience with other customers with similar behavioral patterns. However, usually only transaction data with the company in question are included in the database, while relevant marketing data, for example, on the use of competitive products, are lacking and need to be collected in separate surveys among a sample of customers. In addition, the development of techniques for the extraction of relevant information from the database for strategic marketing purposes, often referred to as data-mining, has lagged behind the development of tools for collecting and storing the data.

In this study, we develop a new data-augmentation tool to predict consumption of new or current products by current customers who do not use them yet. We provide a mixed data factor analyzer that is tailored to implement cross-selling based on customer transaction data and identifies the best prospects for each service. The model extends previous factor analysis procedures and enables us (1) to analyze data from a variety of different types, i.e. choices, counts, or ratings; (2) to represent the variability of those variables in a latent subspace of reduced dimensionality; and (3) to analyze data from the customer database in combination with survey data collected only on a sample from the customer database. The main purpose in applying the model is to learn from the behavioral patterns of all customers in the database and from external data gathered from a survey of a sample of customers, to identify the best prospects for the cross-selling of services, so that each customer is only offered a service she is very likely to be interested in.

The remainder of this paper is organized as follows. In Section 2, we provide a framework describing the role of cross-selling as a tool to enhance customer relationships and review relevant literature on cross-selling. Then, we explain a new mixed data factor analyzer to identify cross-selling opportunities from customer transaction databases. We show how it extends recent work on factor analysis for non-normal variables. Next, the model is calibrated on a customer transaction database from a large retail bank. We compare our model to alternative models and investigate which has better performance in evaluating ownership of financial services. Finally, we discuss other potential applications as well as limitations.

2. Cross-selling

Cross-selling pertains to efforts to increase the number of products or services that a customer uses
within a firm. Cross-selling products and services to current customers has lower associated cost than acquiring new customers, because the firm already has some relationship with the customer. A proper implementation of cross-selling can only be achieved if there is an information infrastructure that allows managers to offer customers products and services that tap into their needs, but have not been sold to them yet.

Furthermore, we conjecture that cross-selling is effective for customer retention by increasing switching costs and enhancing customer loyalty, thus directly contributing to customer profitability and lifetime value. The more services a customer uses with the firm, the higher the costs of switching to other firms, which leads to loyalty and tenure. We illustrate this in Fig. 1. The graph is derived from the empirical application below and shows the number of years of being a customer versus the number of services used. Fig. 1 reveals a strong positive relationship of the number of years of being a customer and the number of services used from the bank. Although causality cannot be demonstrated, there is likely a mutually reinforcing effect. As the length of the relationship increases, customers are inclined to use more services from the bank and, when more services are used, switching costs increase, so that ending the relationship with the bank becomes less attractive. Thus, customer retention is enhanced through cross-selling as switching costs increase with multiple service relationships.

As the intensity of satisfactory interaction with the customer increases, the firm learns more about the customer’s needs and wants, increasing its ability to develop customer loyalty and fend-off competitors. At the same time, the enhanced loyalty leads to increased profitability. Therefore, use of more services leads to higher profits, if the services are properly cross-sold. We illustrate this in Fig. 2, again derived from our empirical data set described below. This figure plots the profitability of a customer against the number of services s/he uses from the bank. One can see again that there is a significant positive relationship, showing that cross-selling directly generates increased profitability by enhancing the life-time value of customers.

Fig. 1. Number of years of using the bank plotted against the number of services used, with 95% confidence intervals.
Despite its importance for relationship marketing, cross-selling has received limited attention in the academic literature. Most of the literature focuses on methodology for identifying common acquisition patterns of products by customers based on their usage or ownership data. The problem is to infer the longitudinal pattern of acquisition across various products or services, when only cross-sectional data are available on usage or ownership. One of the earliest attempts is the study by Paroush (1965), who uses Guttman’s (1950) coefficient of reproducibility as an indicator of the order of acquisition implied by cross-sectional data. Paroush’s study has been replicated and extended by Hebden and Pickering (1974), Kasulis, Lusch, and Stafford (1979), and Stafford, Kasulis, and Lusch (1982).

However, the models used in these studies were not explicitly developed to implement cross-selling. Kamakura et al. (1991) propose a uni-dimensional latent-trait model that makes probabilistic predictions that a consumer would use a particular product or service, based on their ownership of other products/services and on the characteristic of the new one. They apply this latent-trait model to survey data on the use of financial services. However, the approach requires that the firm knows about each customer’s usage of services from both the firm and its competitors, something unlikely to be observed in practice. In most cases, information on ownership of competitive products is available only when collected as a sample of a firm’s customers. Such incomplete data cannot be analyzed with the model of Kamakura et al. Moreover, their specification is limited since it assumes that a single unobserved dimension adequately summarizes the variation of the variables contained in the transaction database and it can only handle binary (0/1) variables, whereas transaction databases usually contain a wide variety of different variables, such as counts, choices, ranks, and classifications.

To accommodate these requirements for a parsimonious model for the description of cross-buying and its use for cross-selling purposes, we extend the recent literature on factor analysis for non-normal variables and exploit its strengths in the imputation of missing data. Our approach builds on recent work in factor analysis for non-normal variables, in particular that by Bartholomew and Knott (1999), Kamakura and Wedel (2000), Moustaki and Knott (2001), and Wedel and
Kamakura (2001). We extend that work in two ways. First, by developing a factor analyzer for mixed outcome data, simultaneously dealing with missing observations. Previous work in this area, as cited above, has not accommodated such mixed outcome data, where some variables pertain to choices, others to ratings, some others to rank-ordered variables, and others to counts. Such a mix of data types is fairly typical in customer transaction databases and its proper analysis is a non-trivial exercise. It is important to accommodate the measurement scales of the variables in forecasting the success of cross selling efforts, where predictions need to be confined to the proper support. A second extension of past work on factor analysis is that we deal with missing data that arise due to sub-sampling. Again, this situation arises fairly often in customer transaction databases, where the transaction data is augmented with a survey among its customers. In addition, the approach that we propose next offers advantages over the one that has been postulated by Kamakura et al. (1991) in that it accommodates a much broader range of distributions of observed variables, allows for multiple dimensions, and allows for predictions that extend beyond the information available within a firm’s customer database.

3. A mixed data factor analyzer for identifying cross-selling prospects

Customer-oriented businesses have a wealth of customer information at their disposal, generated from their data production systems. Harnessing this rich source of customer level transaction information is increasingly important to marketers. Database marketing (DBM) involves building, organizing, supplementing, and mining customer transaction databases to increase the accuracy of marketing efforts by enabling the identification of the best prospects for marketing efforts (Goodman, 1992; Labe, 1994). Many DBM efforts have been ineffective, however, since the database is only used as a mailing list and the possibilities for integration of marketing and computer systems are not effectively exploited (Shaw, 1993). Two causes of this undesirable state of affairs can be identified. First, in many cases, detailed transaction data pertaining to the company in question are compiled, possibly enriched with ZIP-level Geo-Demographic data, but critical data on the use of products and services from competitors, and “soft data” such as customer satisfaction, are lacking. These often need to be collected in separate surveys. Due to the survey costs, such data are usually only collected from a sample of customers in the database. Yet, this type of information is needed for all customers for the effective implementation of one-to-one marketing. Second, the development of methods for the extraction of information for strategic marketing purposes has lagged behind the development of techniques for the construction and maintenance of the databases. Too few efforts have been made to tailor these methods to optimally match the structure of the database or the substantive marketing problem.

To effectively cross-sell its products/services, the marketer must find dependencies among product/service ownership, i.e. must identify the structure in customers’ cross-buying behavior. In particular, one is interested in the likelihood that a particular customer will buy certain products or services that s/he does not own yet, given ownership of other products and services. We develop next a mixed data factor analyzer that is tailored to analyze cross-buying for the implementation of cross-selling based on customer transaction data and identifies the best prospects for each service.

3.1. Description of the factor analyzer

We assume that a firm has access to a customer transaction database and has conducted a survey among a random sample of its customers. Data from this sample survey serves to supplement the customer database, providing, in particular, information about usage of services from competitors. Thus, for a representative sample of its customers, the firm has complete information. Let \( n = 1, \ldots, N \) denote customers in the database and \( j = 1, \ldots, J \) represent observed variables. These \( J \) variables are measured on a variety of scales. In the application below, for example, income and education are rated on ordinal scales, volume of customer transactions on a ratio-scale, the total number of transactions is a discrete count, and service usage is measured with binary indicators. We assume the \( J \) observations, \( y = (y_n) \), to be realizations of random variables, distributed in the exponential
family of distributions. The exponential family is a very general class of distributions, including both continuous and discrete distributions, which allows us to accommodate the various types of data typically encountered in DBM in a single framework, by assigning each observed variable \( j \) its own distribution. For example, binary indicators of service usage can best be modeled with a Bernoulli distribution, numbers of transactions with a Poisson distribution, rating scales with a rank-order binomial distribution, and the volume of transactions with a normal distribution. The exponential family allows one to optimally match the support of the selected distribution to the assumed measurement scale of the transaction variables. This is particularly important in predicting service usage for cross-selling, since individual-level predictions need to be logically consistent with each variable’s measurement scale.

We aim at identifying a low-dimensional map of the observed variables that identifies the most salient features of these data and allows for graphical representation. \( X \) is the \((N \times P)\) stochastic matrix representing that low \((P)\)-dimensional space, where we assume that the elements of \( X \) are independently distributed across subjects according to a standard normal distribution. We specify the conditional distribution of the observations for one particular subject:

\[
f(y_n | x_n) = \prod_{j=1}^J \exp \left[ \frac{y_{nj} \eta_{nj} - a_j(\eta_{nj})}{\phi_j} + b_j(y_{nj}, \eta_j) \right],
\]

\[\eta_n = \lambda_0 + x_n \mathcal{N}.
\]

Here, \( y_n = (y_{nj}) \) is a vector of observed data from customer \( n \), \( x_n \) is the \( n \)-th row of an unobserved vector of i.i.d. normally distributed \((N \times P)\) quantities \( X \), \( \Lambda \) a \((J \times P)\) matrix, and \( \lambda_0 \) a \((J \times 1)\) vector of fixed, but unknown, weights, is a dispersion parameter that applies for certain distributions in the exponential family such as the normal, \( a()\) and \( b()\) functions depending on the particular distribution for the variable \( j \) (McCullagh & Nelder, 1989). Eq. (2) shows that the expectation of the observation vector for each subject is mapped onto a lower-dimensional subspace: \( \eta(x_n) \) defining that map. Note that the specification of the distributions in Eq. (1) implies: \( \mathbb{E}[y_n | x_n] = h(\eta(x_n)) \), with \( h() \) a canonical function, depending on the distribution of the data (it is the log-function for the Poisson and the logit-function for the binomial distribution, for example). Also note that the \( J \) observations on each individual, \( y_{nj} \), are conditionally (but not marginally) independent, given \( x_n \). Since \( x_n \) is normally distributed, so is \( \eta(x_n) \).

Our model provides a factor analyzer, since the reduced \( P \)-dimensional space spanned by captures the salient features of the data and lends itself to a graphical representation of the weights that define the map. We specify the subject-specific map to have a prior normal distribution across subjects: \( x_n \sim \mathcal{N}(0,1) \). The use of the standard normal distribution for the latent variables alleviates scale and translation invariance of the model. Those arise because one can add a vector of scalars to \( x_n \) and subtract a vector of constants from \( \lambda_0 \), or one can post-multiply \( x_n \) and \( \Lambda \) with the inverse of a diagonal matrix \( T \), which yields the same model, as in standard factor analysis.

The factor analyzer provided in Eqs. (1) and (2) is a powerful approach, since it maps observed variables of a wide variety of measurement scales nonlinearly onto a latent feature space of reduced dimension that lends itself for identification of important aspects of the data through graphical display. We view our model as one allowing for convenient graphical display of the structure of data, without a necessary interpretation of the factors as “latent dimensions”. While we see our approach as useful for data reduction and data-mining, similar to PCA, we think that one should be careful in interpreting the results of factor analyses of behavioral data as latent perceptions or intentions. The reason is that, in making inferences on latent dimensions extracted from measurements of behavior, one makes strong claims with respect to the underlying process. Thus, contrary to the application of factor analysis to the analysis of measurement scales specifically designed for the identification of latent dimensions, the application of our tool to customer transaction databases is one where one is not primarily interested in a behavioral interpretation of the latent dimensions, but rather in a convenient low-dimensional graphical display of the structure in the data. However, the maps themselves are interpretable as we will show below.
Note that the distribution in Eq. (1) presents the conditional distribution of the data $Y_i$ given the latent variables $X_j$. To illustrate the form of the expression, assume that there are $J = J_1 + J_2 + J_3$ variables, with, respectively, a normal, Bernoulli, Poisson, and rank-order binomial distribution, as in the application below. Then the conditional distribution of the observed data given latent variables takes the following form:

$$f(y_n | \eta(x_n)) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[\frac{\sigma^{-2}(y_{nj} - \eta_{nj})^2}{2}\right] \times \prod_{j=J_1+1}^{J_1+J_2} \frac{\exp[y_{nj}\eta_{nj}]}{1 + \exp[\eta_{nj}]} \times \prod_{j=J_1+J_2+1}^{J_1+J_2+J_3} \frac{\exp[y_{nj}\eta_{nj} - \exp[\eta_{nj}]]}{y_{nj}!} \times \prod_{j=J_1+J_2+J_3+1}^{J} \binom{K_j - 1}{y_{nj} - 1} \times \frac{\exp[(y_{nj} - 1)\eta_{nj}]}{(1 + \exp[\eta_{nj}])^{K_j - 1}}. \tag{3}$$

Here, $\eta_{nj} = \lambda_0 + x_n \lambda_j$, where $\lambda_j$ is the $j$-th row of $\Lambda$, and $K_j$ is the number of scale points of rank-order rating scale $j$.

It is of interest that our model requires the conditional distribution of the data, given the factor scores, to be in the exponential family. However, since the factor scores themselves follow a normal distribution, the marginal distribution of the data—obtained by integrating over the factor score distribution—is in general not in the exponential family and will accommodate overdispersion. In addition, our model assumes the observed variables to be conditionally independent, given the factor scores. However, our model accommodates marginal dependence of the variables, since they depend on the same unobserved factor scores. We consider these important features in modeling marketing data.

### 3.2. Estimation using SML

The unconditional distribution of the observations is obtained by integrating out the unobserved variables in Eq. (3). The likelihood of the factor analyzer, providing the support of the data for the parameters, is obtained as the product of that expression over all $N$ observations. However, in applications to cross-selling, the observation vector is complete only for a sample of the subjects in the database, being obtained both from the database and the supplementary survey. For the remaining customers in the database, part of the data is missing and, for those subjects, we partition the observation vector as $y_n = (y_n', \hat{y}_n)$, with the corresponding sets of variables being $C = C \cap \hat{C}$, where we assume the first subset of variables to be observed without loss of generality. Also, we assume the customers to be ordered such that for the first $M$ subjects complete data are available, while for the remaining $N-M$ subjects the data are incomplete. The observed data likelihood is obtained by integrating the joint distribution of the observed and missing data over the distribution of the missing data in the likelihood:

$$L(\Xi | Y) = \prod_{n=1}^{N} \int \int \prod_{j \in \hat{C}} f(\hat{y}_{nj} | \eta(x_n), \Xi) \times \prod_{j \in \hat{C}} f(\hat{y}_{nj} | \eta(x_n), \Xi) f(x_n) dx_n, \tag{4}$$

where we collect all parameters in $\Xi$. However, since the data are missing at random (MAR), the survey being conducted among a random sample of the database, this expression is equivalent to the simpler observed data likelihood:

$$L(\Xi | \hat{Y}) = \prod_{n=1}^{M} \int \prod_{j=1}^{J} f(\hat{y}_{nj} | \eta(x_n), \Xi) f(x_n) dx_n. \tag{5}$$

Note that, in Eq. (5), we may ignore the missing data generating mechanism and replace the product over $N$ (all subjects in the database) by a product over $M$ (all subjects in the sample). We may ignore the missing data generating mechanism and use only complete data because the missing data are MAR, being under control of the researcher, the estimators based on Eq. (5) being unbiased (Little & Rubin, 1987).

The estimation of the factor analyzer is not feasible with standard (numerical) algorithms for maximizing the likelihood function, given the potentially high-dimensional integration involved in the likelihood. However, simulated likelihood (SML) estimation has
made the approximation of such integrals possible. Such simulation methods were introduced by McFadden (1989) and an overview is provided by Stern (1997). The problem is to evaluate the log-likelihood (Eq. (5) in the general case where \( x_n \) is a \( P \)-dimensional normal random variable. The idea of simulation is to draw \( S \) random variables \( z_n^S \) from \( f(x_n) \) and use the approximation:

\[
\hat{L}(\xi | \hat{y}) = \sum_{n=1}^{M} \ln \sum_{j=1}^{S} \prod_{j=1}^{J} f(\hat{y}_{nj} | \eta(z_n^j); \xi) / S
\]  

(6)

instead of Eq. (5). The value of \( \xi \) that maximizes Eq. (6) is the SML estimator. SML provides consistent estimators if \( S \to \infty \) as \( M \to \infty \). Then the simulated likelihood (6) is a consistent simulator of the likelihood (5). The bias in the estimates is of order \( 1/S \). However, finite values of \( S \) are sufficient to obtain good properties of the estimates. We use \( S = 100 \) (Lee, 1995).

3.3. Model selection and prediction

In most applications of the factor analyzer, the number of dimensions \( P \) is treated as unknown and needs to be determined empirically. Models with different numbers of factors cannot be compared using standard likelihood-based tests, since the asymptotic \( \chi^2 \) distribution of the LR test of the \( P \)-factor model versus the \( P + 1 \)-factor model does not hold (Anderson, 1980). In order to determine the number of latent factors, we compare the solutions with different numbers of factors on the basis of the consistent Akaike information criterion (CAIC) (Bozdogan, 1987) and choose the solution with the lowest CAIC.

In order to predict/impute the missing data for all subjects in the transaction database, we compute the posterior expectation of these missing data, given the model estimates, and the values of the observed data for the subject in question:

\[
E[\hat{y}_{nj}] = \int \hat{y}_{nj} f(\hat{y}_{nj} | \eta(\hat{x}_n); \hat{\xi}) d\hat{y}_{nj}
\]  

(7)

Here, \( \hat{x}_n \) is a vector with the posterior estimates of the factor scores for customer \( n \); the integrals are again computed through repeated draws from the distributions in question. Currently, our imputations are based on the expected value in Eq. (7), but multiple imputations obtained as draws from the predictive distribution of the variable in question, with expectation as in Eq. (7), can also be generated (Little & Rubin, 1987).

4. Empirical illustration

4.1. Database marketing in the financial industry

In the US, the recent repeal of the Glass-Steagall Act lead to a wave of mergers in financial markets, blurring the distinction between banks, insurers, and brokerage firms. These mega mergers lower the barriers among financial industries (Shesbunoff, 1999). Conglomerates may capture all aspects of a consumer’s financial needs, from checking accounts to life insurance and one-stop shopping for financial services will become common. On the demand side, consumers want to spend less and less time with a financial service provider; electronic banking and e-trading have reduced the opportunity for personal selling and the Internet has made information search less costly and financial markets more transparent to consumers. These developments have stimulated banks to shift from a product focus to a customer focus. As the cost of acquiring new customers increases, financial institutions are coming to the conclusion that their current customers are by far the best prospects for the sales of current and new services and attempt to consolidate service sales from their customers by implementing customer relationship management. DBM is viewed in banking as one of the most powerful marketing tools, but its success depends on the availability of databases (Onge, 1999). The level of penetration of electronic banking has propelled electronic storage of customer transactions, which is now routine in the entire financial industry. Therefore, the financial services industry presents all conditions to the successful implementation of DBM.

4.2. Internal and external data

In order to illustrate the proposed approach for the cross-selling of services, we apply it to a sample of 5550 customers of a major commercial bank in Brazil.
For each of these sampled customers, we have data that were gathered in a personal interview, as well as transaction data from the bank’s internal records. For this particular study, we use the following variables from the bank’s internal records (assumed distribution in parenthesis):

- Number of transactions/month (Poisson)
- Volume of deposits in the bank (normal)
- Education (rank-order binomial)
- Age (rank-order binomial)
- Gender (Bernoulli)
- Ownership of automobile, telephone, fax, and personal computer (Bernoulli)
- Personal income (rank-order binomial)
- Usage indicators for 22 financial services within the bank (Bernoulli). These include four types of services:
  - Conveniences: ATM card, phone banking, PC banking, safety box, private manager, and automatic bill payment
  - Investments: special checking, savings, certificate of deposit, mutual fund, annuities fund, investment fund, commodities fund, and gold
  - Risk management: life insurance, car insurance, and homeowner’s insurance
  - Credit: mortgage, installment loan, credit card, personal loan, and farming credit

These internal data are supplemented with survey data on each customer’s usage of the same 22 financial services from competing vendors (Bernoulli). Note that most of these financial services can be owned from multiple banks by the same customer. Table 1 provides a summary description of the variables in the study.

For this application, we use the complete data on the sample of 5550 customers, since we want to validate our procedure. We estimate the proposed mixed data factor analyzer on a sample of 1387 of these customers. This sample is a random sample taken from all customers and is representative of the entire database and is large enough for reliable estimation of the parameters of our model in reasonable computation times. We then apply the estimated model to the remaining 4163 customers, for whom we assume the survey data on competitive ownership to be missing. We thus predict their likelihood to use each of the 22 services from competing firms, based solely on these customers’ internal records. This is an important problem for the bank in itself, since competitive ownership is only known for a subset of its customers and our procedure allows one to forecast it for all customers in the database. Since in our application we have the survey data for the hold-out customers as well, this allows us to investigate the performance of the procedure, by comparing the imputed values to the “true” values of the survey variables. Our objective is to demonstrate that, once the model is estimated on a combination of internal and external data for a sub sample, it can be applied to the firm’s entire customer database to predict whether customers satisfy their needs for specific financial services elsewhere.

4.3. Results

Estimation of the factor model to the data from both sources leads us to choose the model with three factors ($P = 1$: CAIC = 81,899, $P = 2$: CAIC = 76,571, $P = 3$: CAIC = 75,622, $P = 4$: CAIC = 75,872). In addition to this model, we also estimated the latent-trait model previously proposed by Kamakura et al. (1991), as well as a three-dimensional binary data factor model (Bartholomew & Knott, 1999) for comparison. Note that our application of the latent trait model proposed by Kamakura et al. is an extension of their approach, because we augment the customer database with external (survey) data and utilize the model to make predictions about usage of the services from competing vendors.

The factor weights representing the reduced space map are graphically displayed as vector termini in Figs. 3 and 4, classified by type of service (credit, investment, risk management, or convenience services). Since pictorial information is more quickly processed and better remembered than verbal or numerical information (Spence & Lewandowski, 1990), the graphs allow for efficient communication with bank managers, who can quickly grasp the dependencies of service usage and identify implied cross-selling opportunities. The graphical display facilitates dissemination of the results within the company. Thus, we emphasize low-dimensional representation of the data and graphical display, rather than substantive interpretation of the factors themselves.
The plots allow managers to quickly identify effective strategies for cross-selling, i.e. they enable managers to target services to customers who currently use them from competitors or have a high predicted probability of usage, but have not yet acquired the service within or outside the bank. Those services are identified from the similarities in the weights of the internal and external service items in Figs. 3 and 4, respectively. For example, the fact that the vectors representing credit card (crdt crd) and personal loans (prsnl loan) within the bank are close
to each other indicates that a customer who uses personal loans with the bank would be a good prospect for the bank’s credit card, if she does not use it yet. The same conclusion can be drawn for home insurance (home insr) and car insurance (car insr), indicating cross-selling opportunities.

Similarly, the fact that the vectors representing annuities funds (annuity fnd) within and outside the bank point in the same direction indicates that a customer with a high propensity to use this service might have it from multiple sources. Therefore, s/he would represent a good prospect for strategies that

Fig. 3. Weights for service usage outside the bank.
induce switching (if s/he does not yet use the service within the bank) or a higher “share of wallet” (if s/he already uses the bank’s annuity fund). The opposite conclusion can be drawn about home mortgage (mortg); the maps indicate that usage of this service within the bank is unrelated to usage at another institution. These patterns of joint usage of these services within and outside the bank can be potentially useful to develop a cross-selling program. However, for such a program to be effectively implemented, the question arises to what extent the model makes use of the inter-relationships among these
services to produce accurate predictions, which we consider next.

Fig. 5 shows how the latent space relates to customer demographics. Comparing this figure with the previous ones, one can see that usage of financial services in general is correlated to the demographic variables that indicate income (income, education, ownership of durables, etc.). As one would expect, ownership of information-technology durables (PC and fax) is highly collinear with education and volume of deposits is highly collinear with the number of transactions per month.

4.4. Out of sample tests and comparison

As a validation test, we simulate its application in the identification of customers who use various financial services at competing financial institutions. With this particular purpose in mind, we compute the predictive probabilities that each customer uses the financial services outside the bank for each of the 4163 customers in our hold-out sample, using only information from the internal records as in Eq. (7). Again, this is an important problem for banks since competitive ownership tends to be available only for a sample of the customers based on a survey and our procedure allows one to predict it for all subjects in the transaction database. Since in our particular application we also have the survey data available for the hold-out customers, this enables us to validate the predictive performance out of sample using the actual information obtained from the survey.

As a measure of selectivity of the predictive model, we generate power curves of the cumulative proportion of actual users of the service (observed data) against the cumulative proportion of customers with a certain predicted usage probability for that service (out of sample forecasts). These power curves are shown in Fig. 6, comparing the performance of the three models for selected financial services from the database. The plot for the private manager service, for example, shows that the 30% of customers that are predicted by the three-factor binary model to have the highest probability (based solely on service usage data within the bank), account for more than 70% of all users of that service outside the bank. For our proposed model, that number is 80%. For most services, the power curve for our model lies above that for the
Fig. 6. Power curves for service usage outside the bank.
Fig. 6 (continued).
Fig. 6 (continued).
two competing models. We note that large improvements are attained for cutoffs around 40–50%. This means that the improvement of our model over the two competing models increases substantially if cross-selling is considered for larger proportions of the database. Note that this holds for any predictive model because, as the penetration for a product/service reaches saturation, there is not much to gain in using a more discriminating predictive model.

As a summary measure of selectivity, we compute the Gini coefficient from the power curve for each of the services. The Gini coefficient is a measure of concentration, indicating the extent to which usage of the service outside the bank is concentrated among those customers who were predicted to have a high probability of doing so. We compute this index for service \( j \) as \( G(j) = \sum_{n} v_{nj} - \tilde{v}_{nj} / \sum_{n} 1 - \tilde{v}_{nj} \), where \( \tilde{v}_{nj} \) is the proportion of the sample of customers who have a predicted probability of usage for the service equal or greater than customer \( n \)'s, and \( v_{nj} \) is the proportion of actual users of the service who are ranked equal or higher than customer \( n \) in their usage probability. This measure of concentration equals the ratio of the area between the power curve and the 45° line over the total area above this line. An index equal to zero indicates a lack of predictive power, while a value of one is obtained if the model sorts all customers perfectly in decreasing order of true likelihood of usage. Table 2 shows the Gini coefficients for all 22 financial services. The table reveals that our model yields a substantially higher Gini coefficient than the \( P=1 \) latent trait model in all cases, but one. It yields a higher Gini coefficient than the \( P=3 \) latent trait model in all cases, but four. Thus, the predictive performance of our model is superior to that of the two competing models (note also that the application of the three-dimensional binary factor model to DBM and imputation of missing observations with each of the three models is itself new and that we had to extend the two competing procedures to deal with the missing data structure of the application to enable comparison). In absolute terms, the Gini coefficient is very high (\( G(j)>0.6 \)) for at least six services (safety box, PC banking, farming credit, mutual fund, private manager, and gold) and reasonably high for at least six others (\( G(j)>0.4 \)). These services appear from our analyses to be important for developing cross-selling activities.

**4.5. Identifying the best prospects for cross-selling**

The best prospects for the cross-selling of a particular service are those customers who have a high predicted propensity to use the service within the bank, but do not yet use it. However, these customers may already use the (same) service at a competing financial institution and, therefore, must be persuaded to switch service providers. While the bank does not have perfect information as to whether these customers use the service at a competing institution (it has that information only for the sample of subjects included in the survey), it can use our factor analyzer to compute their propensity to do so and the predicted probability. Fig. 7 illustrates this for four different services, showing the predicted probabilities of usage within and outside the bank, among all customers who do not use the service within the bank, ranked in decreasing order of their potential as cross-selling prospects. For example, Fig. 7 shows that the top 30% prospects for the cross-selling of **phone banking cards** have predicted usage probabilities that are equal or greater than 80%. On the other hand, these same customers have very similar probabilities of being current users of the service at a competing institution. The situation is

<table>
<thead>
<tr>
<th>Service</th>
<th>( P=1 ), binary</th>
<th>( P=3 ), binary</th>
<th>Full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings</td>
<td>0.11</td>
<td>0.13</td>
<td>0.18</td>
</tr>
<tr>
<td>Credit card</td>
<td>0.22</td>
<td>0.24</td>
<td>0.35</td>
</tr>
<tr>
<td>ATM card</td>
<td>0.15</td>
<td>0.16</td>
<td>0.24</td>
</tr>
<tr>
<td>Phone banking card</td>
<td>0.36</td>
<td>0.47</td>
<td>0.44</td>
</tr>
<tr>
<td>CD</td>
<td>0.47</td>
<td>0.55</td>
<td>0.52</td>
</tr>
<tr>
<td>Special checking</td>
<td>0.24</td>
<td>0.25</td>
<td>0.34</td>
</tr>
<tr>
<td>Safety box</td>
<td>0.64</td>
<td>0.66</td>
<td>0.73</td>
</tr>
<tr>
<td>PC banking</td>
<td>0.48</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>Auto bill payment</td>
<td>0.28</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>Personal loans</td>
<td>0.21</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>Mortgage</td>
<td>0.15</td>
<td>0.16</td>
<td>0.26</td>
</tr>
<tr>
<td>Installment loan</td>
<td>0.14</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Farming credit</td>
<td>0.56</td>
<td>0.55</td>
<td>0.63</td>
</tr>
<tr>
<td>Mutual fund</td>
<td>0.60</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td>Investment fund</td>
<td>0.30</td>
<td>0.32</td>
<td>0.39</td>
</tr>
<tr>
<td>Commodities fund</td>
<td>0.41</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>Annuities fund</td>
<td>0.48</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>Private manager</td>
<td>0.49</td>
<td>0.55</td>
<td>0.66</td>
</tr>
<tr>
<td>Gold</td>
<td>0.72</td>
<td>0.72</td>
<td>0.68</td>
</tr>
<tr>
<td>Car insurance</td>
<td>0.34</td>
<td>0.38</td>
<td>0.40</td>
</tr>
<tr>
<td>Home insurance</td>
<td>0.35</td>
<td>0.41</td>
<td>0.43</td>
</tr>
<tr>
<td>Life insurance</td>
<td>0.30</td>
<td>0.38</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Fig. 7. Usage probabilities among non-using customers.
more severe for automatic bill payment. For this service, the top 30% prospects for cross-selling have predicted usage probabilities of only 30% or greater. Moreover, most customers have a higher probability of being current users at a competing institution than to use this service within the bank. Therefore, the selection of prospects for cross-selling purposes will depend on whether the bank can persuade these customers to switch away from competitors.

5. Summary and conclusions

The information revolution in the past couple of decades has caused a proliferation of customer databases, often leading to injudicious applications of direct marketing techniques, canvassing the market with ineffective sales pitches, increasing consumer resistance to “junk mail” and telemarketing, and reducing the profitability of marketing activity. Appropriate use of DBM enables firms to effectively leverage on knowledge about current customers. This maximizes the yield of the sales effort, minimizes the risk of annoying the customer with uninteresting offers, and strengthens the ties between the firm and the customer. As a consequence of the information revolution, in particular in the financial sector, firms have amassed vast amounts of behavioral and demographic data about their customers in data warehouses. Effectively utilizing this source of information requires the application of methods concisely tailored to the requirements posed by the structure and contents of the database and the marketing needs of the company.

In this study, we have presented a novel tool to support cross-selling with database marketing that we believe meets those needs. The method is tailored to a situation where the transaction database is augmented with information on ownership of products and services from competitors, collected through a survey. Our mixed data factor analyzer allows the firm to predict its customers’ likely buying behavior beyond the products and services currently owned from the firm. These probabilistic predictions form the basis for selecting the best prospects for the cross-selling of new products or services. The flexibility of the approach in matching the assumed distribution with the measurement scale of the observed variables is particularly important for cross-selling predictions to be logically consistent. A major advantage of the factor analyzer for strategic use within companies is that its use and implementation is almost entirely based on graphical representation of the dependencies in the data. The identification of services for effective cross-selling and the selection of cross-selling targets can all be accomplished by graphical means, which greatly facilitates ease and speed of use by managers.

The estimation of the proposed model on a random sample combining data from the customer database and from external sources is computationally intensive, due to the reliance on simulation-based estimation methods. Calibration of the model on complete transaction databases of millions or even several hundreds of thousands of customers is currently not feasible. With current technology, it is feasible to estimate the model on sample sizes in the order of tens of thousands, which are typical of large-scale customer surveys. However, the model need not be calibrated on the entire database but only on a sample from it, which, as we have shown, yields reliable estimates of the model parameters. Once the model is calibrated on the sample, the implementation on the entire database is relatively fast and easy. In other words, scalability is not a main concern in the implementation of the proposed model, as long as it can be calibrated on a sample of the customer database.

References


How cannibalistic is the Internet channel? A study of the newspaper industry in the United Kingdom and The Netherlands

Barbara Deleersnydera,*, Inge Geyskensb,1, Katrijn Gielensc,2, Marnik G. Dekimpea,c,3

a Department of Applied Economics, Catholic University of Leuven, Naamsestraat 69, 3000 Leuven, Belgium
b Department of Marketing, Tilburg University, Warandelaan 2, 5000 LE Tilburg, The Netherlands
c Department of Marketing Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Received 21 December 2001; received in revised form 2 April 2002; accepted 3 April 2002

Abstract

During the past decade, irrational exuberance has turned into a possibly equally irrational pessimism about what the Internet can accomplish. The fear of getting ruined through cannibalization losses has recently deterred many firms from deploying the Internet as a distribution channel. But do Internet channels really cannibalize firms’ entrenched channels, or is this widely held assumption exaggerated? To answer this question, we apply recent structural-break time-series econometrics to quantify the impact of an Internet channel addition on the long-run performance evolution of a firm’s established channels. Using a database of 85 Internet channel additions over the last 10 years in the newspaper industries of the UK and The Netherlands, we find that the often-cited cannibalization fears have, at least in this information-goods industry, been largely overstated. The Internet therefore need not be disruptive to established companies and channels. This does not, however, imply that firms enjoy free play in setting up Internet channels. In cases where the newly established Internet channel too closely mimics the entrenched channels, substantial cannibalization is more likely to take place.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Internet channel; Cannibalization; Structural-break time-series analysis

1. Introduction

For decades, the increasingly diverse needs of an ever more fragmented market have compelled firms to increase their product variety. Recently, firms are also turning to a second strategy to better address these diverse consumer needs: they increase their channel variety. Both strategies are risky, however, in that a new product or channel may cannibalize the sales of existing products and channels. While many marketers consider cannibalization when assessing the (potential) success of new product introductions (e.g., Mason & Milne, 1994; Reddy, Holak, & Bhat, 1994), there is hardly any knowledge about channel cannibalization. This may explain why Frazier (1999, p. 232) recently observed that “the ultimate effect [of increasing channel variety] on . . . product sales is . . .
unclear”. This lack of research on channel cannibalization is surprising, especially since the issue has become so pertinent through the recent frenzy surrounding Internet channel additions. Specifically, absent scientific research, and after the first Internet euphoria settled down, practitioners now “widely assume that the Internet is cannibalistic [and] will replace all conventional ways of doing business” (Porter, 2001, p. 73, italics added).

This fear that the Internet is cannibalistic has rapidly pervaded the trade press, and is shared by both resellers and manufacturers. Resellers take the stance that the manufacturer who disintermediates by going online is stealing sales that are rightfully theirs. They regard the Internet as a disruptive force that will cannibalize their business, even to the extent that they are made obsolete. In turn, resellers’ cannibalization fears have alarmed manufacturers, many of whom now view Internet channels as a significant source of channel conflict that may cause resellers to provide lower support for the manufacturer’s products, or even to discontinue their distribution (Frazier, 1999). In a recent large-scale survey of Belgian firms, Konings and Roodhooft (2000) find that of all firms that have access to the Internet, only 15% use that site as an additional channel to actually sell products online. In the US, recent estimates indicate that over 40% of all businesses do not yet sell online (http://www.nua.com/surveys), a number that increases to over 70% when excluding the largest businesses (Washington Post, 2001). The fear of upsetting their entrenched channels through cannibalization has deterred many of these firms from deploying the Internet as a distribution channel (Gilbert & Bacheldor, 2000). Other manufacturers have taken their fear one step further. Levi Strauss, for example, already withdrew its Internet channel after 1 year, mainly because of backlash from resellers and out of fear of getting ruined through cannibalization losses (Dugan, 1999).

In sum, many assume that the Internet is cannibalistic and will eventually replace conventional ways of doing business. Is this widely held assumption correct or exaggerated? Moreover, if cannibalization exists, is the downturn in performance of a permanent nature? These questions, in essence, are the focus of the current study. To answer these questions, we propose recent structural-break time-series econometrics to quantify the impact of an Internet channel addition on the long-run performance evolution of the firm’s established channels. Our subsequent empirical application is situated in the newspaper industry. Based on a careful investigation of the cannibalization effects for 85 Internet channel additions over the last 10 years in the British and Dutch newspaper industries, we attempt to provide empirical generalizations as to the magnitude and nature (permanent vs. temporary) of online cannibalization effects. The newspaper industry offers an interesting setting to test the validity of the cannibalization assumption because publishers have taken the lead in exploiting the Internet as a new distribution channel. As such, the publishing industry tends to “act as the pacesetter for the information society” (European Commission, 1996, p. 1), and may foreshadow trends that will occur more slowly in other industries.

2. Literature review

Although there is surprisingly limited empirical work on channel cannibalization, several researchers have expressed their concern about the cannibalization hazards companies face when they add an Internet channel to their entrenched channels (e.g., Alba et al., 1997; Brynjolfsson & Smith, 2000; Coughlan, Anderson, Stern, & El-Ansary, 2001). Both the academic and the trade press typically proclaim cannibalization effects to be present, as is reflected in the following quotes:

Established businesses that over decades have carefully built brands and physical distribution relationships risk damaging all they have created when they pursue commerce in cyberspace. (Ghosh, 1998, Harvard Business Review, p. 126)

Call it survival by suicide. . . . Today’s corporate leaders are building separate new e-enterprises designed to compete head-on with the mother company. (Useem, 1999, Fortune, p. 121)

The fear for online channel cannibalization is nourished by four observations. First, sales may shift from the entrenched channels to the new Internet channel when the latter provides more appealing features to the target audience, such as a quasi-unlimited amount of
information on product attributes, increased customization, time savings, etc. (Alba et al., 1997). Second, the Internet is likely to increase the power of the consumer, as price comparisons across firms can be performed quickly and easily. The resulting increase in price competition may explain why online prices for homogenous products are often found to be lower than those in conventional outlets (Brynjolfsson & Smith, 2000), which may cause sales to shift even further. Third, total sales may also decrease when consumers buy less through the new channel than through their old channel, e.g. when there are less impulse purchases through the Internet (Machlis, 1998). Fourth, existing channels may view the new Internet channel as unwelcome competition. When this happens, the firm’s entrenched channels may lose motivation, and can reduce their support for the firm’s products or even discontinue their distribution. This may, in turn, result in more brand switching towards the firm’s competitors, and hence decreased total sales (Coughlan et al., 2001).

Unfortunately, in spite of the above theoretical observations, and at a time when there is much debate about Internet channel additions in the popular press, empirical evidence on the performance effects of Internet channel additions is still lacking, with the exception of Geyskens, Gielens, and Dekimpe (2002) and Ward and Morganosky (2000). However, the former application, which is also situated in the pace-setting newspaper industry, focuses on the total value implications of Internet channel additions (as measured by stock market returns), while the separate revenue impact of cannibalization is not explicitly considered nor quantified. In the multiple-industry study by Ward and Morganosky (2000), cannibalization is proxied by relating consumers’ self-reported channel usage to information searches in particular product categories. Specifically, consumers were asked to check whether or not they (a) bought and (b) searched for information on the product within the past 6 months from online outlets, catalogs, and/or retail stores. Since online product information searches generated offline sales but not vice versa, the authors conclude that online channels do not cannibalize offline sales. Unfortunately, their cannibalization measure is very crude and makes abstraction of the magnitude of the cannibalization effect. Moreover, the above two studies approach cannibalization solely as a short-run phenomenon, a characteristic they share with the empirical product cannibalization literature. This short-term focus of existing studies is likely to contribute to managerial myopia (Dekimpe & Hanssens, 1999). Clearly, if shortsightedness is to be avoided, empirical studies that derive the long-run impact of (Internet) channel additions are essential. Or, as Frazier (1999, p. 232) pointedly put it: “models need to be developed to help determine when multiple channels need to be relied on [since] the ultimate effect on long-term product sales is [...] unclear” (italics added).

To this extent, we examine whether the long-run evolution of the incumbent channels’ revenues is affected. Specifically, recent advances in time-series econometrics will be used to determine whether the Internet channel addition caused either a persistent growth slowdown or, in the absence of any growth impact, a sustained level drop in the company’s traditional revenue streams.

3. Methodology

Over the last few years, time-series concepts such as unit-root and cointegration tests, error–correction models, and impulse–response and persistence estimates have become popular in the marketing literature to quantify the long-run impact of price promotions (Nijs, Dekimpe, Steenkamp, & Hanssens, 2001; Srinivasan, Popkowski Leszczyc, & Bass, 2000), advertising expenditures (Dekimpe & Hanssens, 1995), and distribution changes (Bronnenberg, Mahajan, & Vanhonacker, 2000). All of these studies, however, look at the overtime performance implications of regular, small shocks, such as an unexpected one-cent price reduction, a one-unit/one-period increase in advertising spending, or a gradual increase in distribution coverage. All these small shocks are assumed not to change the underlying data-generating process (Pesaran & Samiei, 1991).

The marketing event we consider, i.e. the addition of an Internet channel, is not quantifiable in a similar, continuous way. Yet, its impact on performance may well be substantial (Hanssens, Parsons, & Schultz, 2001, p. 293), and can even change the underlying data-generating process. Dekimpe and Hanssens (2000, p. 191), for example, point out that “the
constant parameter assumption in many persistence-based models may no longer be appropriate to test the long-run implications of Internet-related decisions. We therefore extend the aforementioned studies to explicitly account for the fact that the event of interest may represent a different kind of shock to the system, in that it is (a) large, (b) infrequent, and (c) may structurally alter the long-run properties of the time-series (Balke & Fomby, 1991), causing a different steady-state growth (or decline) path to emerge (Ben-David & Papell, 1995, 1998).

Specifically, we start our testing procedure with the structural-break unit-root test, as described in Perron (1989, 1994):

\[ Y_t = \alpha + \beta T + [\theta DU_t + \gamma DT_t + \delta D(TB)_t] \]

\[ + \rho Y_{t-1} + \sum_{i=1}^{k} c_i \Delta Y_{t-i} + \sum_{s=2}^{S} \eta_s SD_{st} + \varepsilon_t \]  

(1)

with \( Y_t \) the log-transform of the performance measure of interest, \( T \) a deterministic trend variable, \( SD_{st} \) a set of seasonal dummy variables, \( \varepsilon_t \) a series of white-noise residuals, and \( DU_t, DT_t \) and \( D(TB)_t \), three structural-break dummy variables, discussed in more detail below. Making abstraction of the terms between brackets, Eq. (1) becomes the well-known Augmented Dickey–Fuller (ADF) test.\(^5\) The lagged first differences are added to ensure that the residual series is indeed white noise, while the parameter \( \rho \) is the key parameter of interest to separate (level or trend) stationary (\( \rho < 1 \)) from evolving, unit-root (\( \rho = 1 \)) series.\(^6\)

We extend the standard ADF test with three dummy variables to allow the exogenous event of interest, i.e. the Internet channel addition, to cause a structural change in the growth rate and/or level of the conventional channel’s performance. As is common in the structural-break literature (see e.g. Ben-David & Papell, 1995, 1998), we allow for a structural change in the trend function of the data-generating process, and do not account for changes in e.g. the model’s autoregressive parameters or error variance. This approach is completely in line with the intervention-analysis approach of Box and Tiao (1975),\(^7\) in that extraordinary events are separated from the regular noise function, and modeled as a change (intervention) in the deterministic part of the time-series model (Perron, 1994). Denoting TB as the potential break date, the dummy variable \( DT_t \) takes the value of \( t - TB + 1 \) when \( t \geq TB \), and zero otherwise. As such, it allows for a change in the growth rate of the trend curve at the time of the channel addition. \( DU_t \) which is a step dummy variable that takes on the value of one when \( t \geq TB \) and zero otherwise, is used to allow for a level shift at the time of the introduction, while \( D(TB)_t \) is a pulse dummy taking the value of one when \( t = TB \) and zero otherwise. This third dummy variable is only added to ensure that the test statistic on \( \rho \) remains invariant in finite samples to the value of the change in the intercept under the null hypothesis of a unit root (Perron, 1994), but has no substantial interpretation in our context.

When \( \rho < 1 \), the unit-root null-hypothesis is rejected and hypotheses tests concerning the other parameters can be carried out using conventional \( t \) and/or \( F \) tests (Holden & Perman, 1994; Perron, 1994). When \( \gamma \) is significant and negative, the introduction of the new Internet channel has a cannibalizing effect in that it initiates a decline (when \( \beta \) was equal to zero) in the performance evolution of the conventional channel, slows down its growth (in case \( \beta > 0 \)), or accelerates an already ongoing decline (for \( \beta < 0 \)). The most dramatic change occurs if \( |\gamma| > |\beta| \) with \( \beta > 0 \) and \( \gamma < 0 \), in which case the sign of the growth path would be reversed after the Internet channel addition, a situation referred to as a growth meltdown (rather than a mere slowdown) by Ben-David and Papell (1998).

The coefficient \( \gamma \), however, measures only the immediate impact on the series’ growth rate. To quan-

\(^4\) This transformation is common practice in time-series analyses, as it allows reducing potential heteroscedasticity in the data. Moreover, the first difference of the log-transformed series provides a good measure for the growth rate of the original series (Franses & Koop, 1998).

\(^5\) See Enders (1995) for a general discussion, and Nijs et al. (2001) for a recent marketing application.

\(^6\) Following Perron (1989, 1994), a general to specific recursive procedure is used for selecting the truncation lag parameter \( k \). Starting with \( k \) equal to 8, it is successively reduced until a model is found where the last lag is significant, while the one-higher lag is no longer significant. If no lags are significant, \( k \) is set to zero. Following common practice (Perron, 1989; Zivot & Andrews, 1992), the criterion for testing the significance on the last lag is set to 1.60.

\(^7\) See e.g., Leone (1987) for a marketing application.
tify the full, long-run growth impact, we rewrite Eq. (1) as:

\[ Y_t = \alpha + \beta T + \theta DU_t + \gamma DT_t + \delta(D(TB))_t \\
+ \sum_{i=1}^{k+1} a_i Y_{t-i} + \sum_{s=2}^{S} \eta_s SD_{st} + \varepsilon_t \tag{2} \]

and compute the difference between the long-run growth \( (\Delta Y_{t,R}) \) of this autoregressive series before and after the Internet channel addition:

\[ \Delta_{L,R, \text{after}} Y - \Delta_{L,R, \text{before}} Y = \frac{\beta + \gamma}{1 - \sum a_i} - \frac{\beta}{1 - \sum a_i} \tag{3} \]

The long-run, or steady-state, growth rate of a series is time-subscript independent, and is obtained as \( \lim_{t \to \infty} \Delta Y_t \). Assume, for ease of exposition, the following first-order autoregressive process with normal starting conditions \( Y_0 = 0 \) and with the error term omitted:

\[ Y_t = \alpha + \beta T + a_1 T_{t-1} \tag{4} \]

In this stylised example, it is easy to show that \( \Delta Y_1, \Delta Y_2, \Delta Y_3, \ldots \) equal \([\alpha + \beta], [a_1 \alpha + (1 + a_1) \beta], [a_1^2 \alpha + (1 + a_1 + a_1^2) \beta], \ldots \) Looking in steady-state, i.e. letting \( t \to \infty \), a long-run growth rate of \( \beta/(1 - a_1) \) is obtained. When working with higher-order autoregressive models, this expression for long-run steady-state growth generalizes to \( \beta/(1 - \sum a_i) \) (see Ben-David & Papell, 1995 for a formal treatment). Similarly, once the trend break dummy \( DT_t \) in Eq. (2) becomes effective, the ‘new’ steady-state growth rate becomes \((\beta + \gamma)/(1 - \sum a_i)\). Standard errors on (the difference in) these steady-state growth rates can subsequently be derived using the well-known delta-method.

When \( \gamma/(1 - \Sigma a_i) \) is not significant, the traditional channel does not experience a long-run growth slowdown because of the Internet channel addition, but its level may be reduced permanently. To assess whether, in the absence of a growth change, cannibalization effects show up in the form of a level drop, we reanalyze Eqs. (1) and (2) without \( DT_t \) and test for the significance of \( \theta \). Depending on the value of the autoregressive parameters \( a_i \) \((i=1, \ldots, k+1)\) in Eq. (2), part of this initial loss may either become persistent, or the initial loss may even be intensified in subsequent periods. Using a similar logic as before, this long-run effect can be shown to equal \( \theta/(1 - \sum a_i) \).

When \( \rho = 1 \), the unit-root null hypothesis is not rejected, causing \( \ell_0 \) and \( \ell_t \) to be no longer asymptotically normally distributed (Holden & Perman, 1994; Perron, 1994). We therefore follow Ben-David and Papell (1998) and Franses (1998) in that we first impose the unit root found through Eq. (1), and estimate the following equation in first differences:

\[ \Delta Y_t = \alpha + [\gamma DU_t + \theta D(TB)]_t + \sum_{i=1}^{k} a_i \Delta Y_{t-i} \\
+ \sum_{s=2}^{S} \eta_s SD_{st} + \varepsilon_t \tag{5} \]

The \( \gamma \) parameter associated with the step dummy \( DU_t \) then gives the immediate impact on the growth rate, while the \( \theta \) parameter associated with the pulse dummy \( D(TB) \), quantifies immediate level shifts (see also Balke & Fomby, 1991). The corresponding long-run effect on the steady-state growth rate can then be shown to equal \( \gamma/(1 - \sum a_i) \) (Ben-David & Papell, 1998). In the absence of any growth impact, we again impose a common slope for the trend function before and after the Internet channel introduction, i.e. we reanalyze Eqs. (1) and (5) without \( DU_t \), and test for the significance of \( \theta/(1 - \sum a_i) \). This parameter allows testing whether cannibalization effects show up in the form of a persistent level drop.

In sum, long-run cannibalization effects will be quantified through the decline in the traditional channels’ steady-state growth rate, or, in the absence of any growth impact, through the size of the persistent reduction in their performance level.

4. Data description

Our empirical application is situated in the newspaper industry. This industry offers an interesting setting to test the validity of the cannibalization assumption for two reasons. First, cannibalization is particularly imminent when information products are delivered online (Shapiro & Varian, 1999). Speed of
delivery, easy search facilities, and customization options tend to increase the value of the information product in the online channel over existing offerings through traditional channels. Second, whereas the newspaper industry shares the above characteristic with other information goods industries, it has the additional cannibalization-prone characteristic that most publishers are still reluctant to charge for their online editions. As such, any switching between conventional and online channels should negatively affect circulation revenue. Moreover, the newspaper industry represents an old and mature industry, where competition is mostly a zero-sum game (Abbring & Van Ours, 1994). Cannibalization fears when going online are therefore even more realistic (see e.g., US Department of Commerce, 1998).

Cannibalization losses may be reflected in a net reduction in either one of the newspaper’s existing income streams: circulation revenue obtained from the number of copies sold and advertising income (Abbring & Van Ours, 1994). With respect to circulation, Mitchell (2001) reports that 50% of all publishers fear that their Internet operations may inflict long-run harm on their print business. Gilbert (2001), on the other hand, argues that, if positioned correctly, print and online versions may actually operate in a complementary way, in that the latter may attract new readers not currently subscribing to the print edition. A similar uncertainty exists on the extent of cannibalization that will arise in the advertising arena. First, newspapers’ advertising incomes are positively correlated with the size of their subscription base, for which the impact is (as argued before) still unclear. Second, Silk, Klein, and Berndt (1999) argue that newspapers are a potentially vulnerable medium to lose their advertisers to the Internet, even though the extent of this loss will again depend on the relative positioning of both media.

Data were collected for 85 online newspaper introductions in two European countries, the UK and The Netherlands. Information on the online introduction dates were gathered by contacting each individual newspaper, and extensively validated through both newspaper archive searches and the Dow Jones Interactive Publication Library. The earliest launch dates were November 1994 and July 1995, whereas the latest launches took place in October 1999 and April 2000 for the UK and The Netherlands, respectively.

Circulation data were collected for 67 daily UK newspapers that went online within the available data period. Circulation is expressed as the average number of copies sold daily. The data are obtained from the ‘Audit Bureau of Circulations’ (ABC). We have 138 monthly observations for 12 national and 55 regional publications (January 1990 until June 2001). On average, 42 (minimum = 20; maximum = 79) observations are available after the introduction date of the online version.

Advertising revenues from the traditional paper edition were obtained from 18 Dutch newspapers that introduced an online version of their paper within the considered time period. The data are expressed in constant guilders and were provided by the ‘Bureau voor Budgetten Controle’ (BBC). For both national (9) and regional newspapers (9), we have 130 monthly observations (January 1990 until October 2000). On average, 41 (minimum = 6; maximum = 63) observations are available after the introduction date of the online version.

5. Empirical results

Cannibalization losses may be reflected in a net reduction in one (or both) of the newspapers’ two existing income streams: circulation and advertising income. We first discuss findings on the circulation data (Section 5.1), followed by the results for advertising (Section 5.2). Section 5.3 provides several robustness checks on our main findings.

5.1. Circulation results

The unit-root null hypothesis is rejected in 14 instances based on the regular ADF test (using the conventional 5% significance level), and in five additional cases when allowing for a structural break at the time of the Internet channel addition. This confirms prior simulation evidence in Perron (1989) that a failure to account for a major exogenous event tends to bias regular unit-tests towards non-rejection of the null hypothesis.

Long-run, steady-state, growth rates are summarized in column 2 of Table 1, and were derived based on the parameter estimates in Eq. (2) for the 19 level/trend stationary series, while Eq. (5) was used for the 48 unit-
root series. Given our interest in the presence of cannibalization, directional (one-sided) tests are used.\textsuperscript{8}

Steady-state growth rates showed a significant decline after the Internet channel introduction in only five series (i.e. a mere 7% of cases). Even though our main interest lies in cannibalization, it is worth noting that in 10 instances, a positive steady-state growth-rate change was observed ($t > 1.64$). Interestingly, the newspaper with the largest positive significant break is the Financial Times, the only English newspaper focusing almost exclusively on economic issues. This finding is in line with Pauwels and Dans (2001), who argue that economically oriented newspapers are likely to experience positive spillover effects from an online channel addition. The average long-run growth impact across all 67 newspapers indicates a small but positive effect, obtained from about an equal number of negative (35) and positive (32) changes that are also approximately equal in absolute value ($/C_0 0.0014$ and $/C_0 0.0018$).\textsuperscript{9}

One could argue that the power of each of the individual tests may not be excessively high. Yet, we concur with Dekimpe, Hanssens, and Silva-Risso (1999, p. 280) that the combined evidence of no growth slowdown in more than 90% of the instances is very strong evidence. On the other hand, one might still object, based on analyses of the parameter sign, that even though there are only 5 significant growth slowdowns, the number of negative changes is quite substantial (35). We therefore conducted a meta-analysis to test for the presence of cannibalization using the one-sided $p$-values associated with the change in the long-run growth rate after the Internet channel introduction, using the method of adding weighted $Z$’s (Rosenthal, 1991). This offers a stronger test for the presence of cannibalization than the significance of the individual impact estimates. Indeed, when the impact on many newspapers is weak (e.g. $p < 0.20$) but in the same direction, case-by-case tests would not reveal any significant cannibalization impact to worry about. Still, the collective evidence, as reflected in the meta-analysis, would suggest a highly significant effect (see Rosenthal, 1991 for technical details, or Dekimpe, Steenkamp, Mellens, & Vanden Abeele, 1997 for a marketing application). However, also the collective evidence showed no support for the presence of a significant cannibalization threat in the newspapers’ circulation-revenue growth ($p = 0.94$).

Cannibalization may not only result in a growth slowdown, however. It may also show up in a level drop. The 52 series where the $|t$-statistic$|$ on $\gamma$ was smaller than 1.64 in Eq. (2) were therefore reanalyzed after imposing a common pre- and post-growth rate (i.e. we deleted $DT_t$ from Eq. (2) and $DU_t$ from Eq. (5)), and we tested for the significance of $\theta/(1 - \Sigma a_t)$ (column 3, Table 1). Four significant negative level shifts were found. When focusing on the average level impact

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
 & Circulation &  & Advertising \\
 & Impact on steady-state growth & Impact on steady-state level & Impact on steady-state growth & Impact on steady-state level \\
\hline
Sample size ($n$) & 67 & 52 & 18 & 16 \\
# negative impact & 35 & 26 & 11 & 7 \\
# significant negative impact (5%) & 5 & 4 & 1 & 0 \\
# positive impact & 32 & 26 & 7 & 9 \\
Average negative impact & $-0.0014$ & $-0.0128$ & $-0.0065$ & $-0.0231$ \\
Average positive impact & $0.0018$ & $0.0130$ & $0.0030$ & $0.0671$ \\
Overall average & $0.0002$ & $0.0001$ & $-0.0030$ & $0.0276$ \\
Rosenthal test on cannibalization & $p = 0.94$ & $p = 0.58$ & $p = 0.13$ & $p = 0.98$ \\
\hline
\end{tabular}
\caption{Long-run performance implications of an Internet channel introduction}
\end{table}
across all 52 series, a pattern quite similar to that of the growth rate change can be observed. Again, a very small average positive effect is found, based on an equal number of negative (26) and positive (26) values that are also almost equal in absolute size (−0.0128 and 0.0130). Moreover, the Rosenthal test did, once more, not reveal a significant level drop \( (p = 0.58) \).

5.2. Advertising results

As with the circulation data, most advertising series are classified as having a unit root: based on the structural-break (regular) test, the unit-root null hypothesis was rejected in 4 (3) instances. Long-run, steady-state, growth rates are summarized in column 4 of Table 1. Of the 18 cases examined, only one experienced a significant reduction in steady-state growth rate \((\gamma/(1 - \Sigma a_i) = -0.0051; \ t = -2.27)\).\(^\text{10}\)

Even though only one newspaper experiences a significant long-run growth slowdown, the number of negative long-run growth-rate changes is substantial (11). Moreover, the average growth impact across all 18 observations is negative (−0.0030). However, as with the circulation data, also the collective evidence from the Rosenthal test did not reveal a significant cannibalization effect \((p = 0.13)\).

The 16 newspapers with a \(|t\text{-statistic}|\) for \(\gamma\) less than 1.64 in Eq. (2), i.e. whose long-run growth was not affected by the Internet channel addition, were subsequently reanalyzed after imposing a common slope. None of the newspapers experienced a significant reduction in their steady-state performance level because of the Internet channel addition (column 5, Table 1). Also the meta-analysis did not reveal such an effect \((p = 0.98)\).

5.3. Robustness checks

Based on this individual and collective evidence, we can conclude that the fear for cannibalization in terms of the newspapers’ circulation and advertising income appears to be overrated. To assess the robustness of this substantive conclusion, we conducted a number of validation checks. Specifically, we (1) assessed the sensitivity of our results to the level of temporal aggregation in the data, (2) tested whether the finding of no cannibalization was driven by the lack of sufficient post-event data, and (3) correlated the growth and/or level changes with the stock-market reactions reported in Geyskens et al. (2002).

5.3.1. Sensitivity to the level of temporal aggregation

Prior research has demonstrated that both unit-root tests (Pierce & Snell, 1995) and inferences about the autoregressive parameters (Rossana & Seater, 1995) may be sensitive to the level of temporal aggregation in the data. Following Nijs et al. (2001), we doubled our temporal aggregation level from a monthly to a bimonthly level.

Our substantive results were not affected. In terms of the circulation data, the same outcome was obtained for the structural-break unit-root tests in more than 80% of the cases. Seven newspapers were found to have a significant negative growth-rate change in the bimonthly series (as opposed to five when working with monthly data), and the collective evidence was again very similar: the average change in steady-state growth is approximately twice as large as for the monthly analysis (−0.0029 vs. −0.0014 for the negative changes, and 0.0036 vs. 0.0018 for the positive changes), while the Rosenthal test, in accordance with the monthly analyses, showed no significant impact, neither for the trend breaks \((p = 0.99)\) nor for the level shifts \((p = 0.89)\).

For the advertising data, none of the unit-root test outcomes was affected. In terms of their long-run growth rate, a significant growth slowdown was found for the same newspaper, with the change in its steady-state growth rate (−0.0089) about twice the change in its monthly growth rate (−0.0051). As before, the average change in growth rate across all 18 observations is approximately twice as large as for the monthly analysis (−0.0111 vs. −0.0065 for the negative changes, and 0.0085 vs. 0.0030 for the positive changes), and the meta-analytic Rosenthal test continued to show no significant impact \((p = 0.24)\). As for the level shifts in advertising income, we again found none of the individual level shifts to be significantly negative, and the meta-analytic test result was again not significant \((p = 0.97)\).
5.3.2. Sensitivity to the timing of the Internet channel introduction

One could argue that for the more recent Internet channel additions, insufficient post-event data points were available to detect cannibalization. As indicated before, on average, 42 post-event circulation observations (minimum = 20; maximum = 79) were available, and 41 post-event advertising observations (minimum = 6; maximum = 63). If the argument above would be valid, however, we would expect more evidence of cannibalization in those series that experienced the Internet channel addition earlier. Two test procedures for this contention were implemented, a regression of the long-run growth-rate/level changes against the introduction time of the Internet edition, and a split-half analysis.11

For the circulation series, we first ran a regression of, respectively, the change in long-run, steady-state, growth rate (using all 67 observations) and of the change in long-run level (based on the 52 observations mentioned earlier) on the time of the Internet channel introduction. No significant relationship was found ($\beta = -2.53 \times 10^{-6}, p = 0.89$ and $\beta = -1.96 \times 10^{-4}, p = 0.36$). Second, the Rosenthal test was conducted on two subsamples, constructed according to a median split on the entry-timing variable. Again, in none of the analyses, any collective evidence of cannibalization was found, neither in terms of a growth slowdown ($p = 0.99$ for early entrants, and $p = 0.49$ for late entrants) nor in terms of a level reduction ($p = 0.72$ for early entrants, and $p = 0.63$ for late entrants).

In terms of the newspapers’ advertising figures, the same two test procedures were applied. The regression against the time of entry did not find any positive relationship between the timing of the Internet channel introduction and the change in either long-run growth (based on 18 observations) or long-run level (based on 16 observations). Rather, the slope coefficient for the long-run growth rate turned out to be negative rather than positive ($\beta = -2.62 \times 10^{-4}, p = 0.014$) and the long-run level shifts showed again no relation ($\beta = 7.39 \times 10^{-5}, p = 0.95$). Similarly, the Rosenthal test on the subsamples of early, respectively, late entrants was never significant, neither in terms of their steady-state growth rate ($p = 0.18$ for early entrants, and $p = 0.26$ for late entrants), nor in terms of their long-run level ($p = 0.69$ for early entrants, and $p = 0.99$ for late entrants).

5.4. Linkage with the stock market reaction to the Internet channel introduction?

Geyskens et al. (2002) argue that Internet channel additions may have value-enhancing (e.g. demand expansion, higher prices, and lower physical-distribution and/or transaction costs), as well as value-destroying (e.g. cannibalization, lower prices, and higher costs) effects. They test their framework in the same industry through a financial event study. Arguing that financial markets are efficient and therefore should reflect all relevant information about these various factors, they report (a) a positive average stock-market reaction around the event date, as reflected in a cumulative average abnormal return (CAAR) around the event day of 0.71%,12 and (b) a mixture of both positive and negative CAARs when looking at the individual newspapers.

Across the two countries considered in our study, 44 newspapers also appeared in the Geyskens et al. (2002) sample. Based on those 44 observations, a significant positive correlation is found between these newspapers’ CAARs and their change in steady-state growth rate (Pearson correlation coefficient = 0.33, $p = 0.029$). Of the newspapers that received a positive (negative) stock-market reaction, 64% (63%) were characterized by a positive (negative) change in their steady-state growth. This result is encouraging, especially when considering that the cannibalization threat is just one of the many value enhancing/destroying factors identified in the conceptual framework of Geyskens et al. (2002), all of which are supposedly reflected in the ultimate stock-market reaction. The relationship between the CAARs and the level shifts is weaker, resulting in an insignificant correlation ($p = 0.41$).

6. Discussion

During the past decade, irrational exuberance has turned into a possibly equally irrational pessimism

---

11 No heteroscedasticity was found using the White test. We therefore applied OLS instead of WLS (see Nijs et al., 2001 for a similar approach).

12 The event window considered was [0,1] days.
about what the Internet can accomplish. The fear of getting ruined through cannibalization losses has recently deterred many firms from deploying the Internet as a distribution channel. Caught up in this general fervor, others have already withdrawn their Internet channels only shortly after having established them. But do Internet channels really cannibalize firms’ entrenched channels? Or is this widely held assumption exaggerated and are practitioners making fundamental mistakes by shying away from this new channel? Unfortunately, this issue of high managerial importance has barely been touched on in empirical research. The time has come to move away from jumping to the presumption that the Internet is automatically cannibalistic and see the Internet for what it is.

Using a database of 85 Internet channel additions over the last 10 years in the British and Dutch newspaper industries, we find that the often-cited cannibalization fear has been largely overstated. Relatively few newspapers in our sample experience a significant drop in their circulation or advertising revenues. This result may imply that there is only little overlap between customers using the traditional channel and those preferring the online channel. Indeed, a number of researchers have recently argued that the readership profile of online newspapers tends to be younger, higher educated, mainly male and more technology oriented than the majority of the print readers (e.g., Nicholson, 2001; Pauwels & Dans, 2001). Similarly, the overwhelming majority of online advertisrs are not print advertisers. Key advertising customers in print are local department stores, grocery chains, and retail clients, whereas most Internet advertising is placed by sellers of computer-related products and services (Gilbert, 2001; Silk, Klein, & Berndt, 2001). Still, this does not imply that firms have free play in setting up Internet channels. An appropriate positioning is still called for, as recently argued by, among others, Gilbert (2001) and Silk et al. (2001). Indeed, when the new channel is positioned as too close a copy (substitute) to its traditional counterpart, cannibalization will more likely take place. Support for this contention is found when comparing newspapers that experienced significant cannibalization losses in their circulation revenues (9) with those that did not (58) in terms of product similarity (overlap) between the online and print version of the product. Data on product similarity were obtained by contacting the newspapers’ webmasters, and asking them to indicate the percentage overlap between the online and print version of the newspaper at the time the Internet edition was first launched, an operationalization also adopted in Gilbert (2001). A t-test on the difference in mean similarity across both groups strongly indicates that newspapers’ circulation revenues are more likely to get cannibalized when there is a high overlap between their online and offline versions (mean similarity is 78% for cannibalized newspapers vs. 45% for non-cannibalized newspapers; \( p = 0.0008 \)).

In sum, our conclusions are encouraging. The threat of cannibalization appears to be considerably lower than widely assumed. The Internet rarely cannibalizes traditional channels. The Internet therefore need not necessarily be disruptive to established companies and channels. Too often, change is confused with disruption. Disruption means invalidating or making substantially less important the advantages of incumbents (Bower & Christensen, 1995). The Internet does not invalidate the importance of entrenched distribution channels. Firms can have both together, and can even create some synergies from a diverse, yet well positioned, channel portfolio.

Limitations and further research. The present study has various limitations, which offer immediate avenues for future research. First, we treated the potential break date as exogenous, i.e. we made full use of prior (managerial) knowledge on the timing of the event. This practice has been criticized as it would allow the researcher to “cherry-pick” the most convenient location of a potential break (see e.g., Christiano, 1992). In our case, however, the managerial information on the timing of the event came from a truly exogenous source, and is objectively known in the market. As such, it was by no means a “convenient”, ex post (e.g. after a visual data inspection) choice on the part of the authors (see Perron, 1994 for an extensive discussion on the topic). Still, one could let the data suggest endogenously the most likely break date (as advocated, for example, in Vogelsang, 1997 or Zivot & Andrews, 1992), and see whether the Internet channel introduction coincides with (or is located closely to) this break point. Second, even though no significant relationship was found between the occurrence of cannibalization and the length of the time period after the event, care should be exerted when extrapolating our findings too.
far outside the sample range. Specifically, our conclusions are based on the assumption that no other, not yet anticipated, technological discontinuity will take place, while the constant-parameter assumption for the autoregressive parameters of the model does not take into consideration that, as multiple years pass by, both online and offline readerships may eventually start to become closer (Silk et al., 2001). Third, there is the issue of external validity. We conjecture that in the absence of cannibalization effects in the cannibalization-prone newspaper industry, the widely claimed cannibalization threat would also appear to be overrated for other, less vulnerable, industries. Yet, we do fully acknowledge that our sample is limited to firms from a single industry. Given that this industry may have some peculiarities, such as the possibility to consume the product (information) in bits and pieces “without destroying” the original physical piece, we strongly suggest for future research to explore to what extent our results are generalizable to other industries. Fourth, the non-availability of data on both revenue streams for the same set of newspapers prevented us from analyzing the interaction that is to be expected between circulation and advertising (Abbring & Van Ours, 1994). Finally, given that cannibalization effects are largely absent, future research could analyze how firms can achieve positive spillover effects from an Internet channel edition, as was obtained in our sample for the more economically oriented newspapers.

Acknowledgements

The second author gratefully acknowledges support from the Netherlands Organisation for Scientific Research (NWO).

References


Do international entry decisions of retail chains matter in the long run?

Katrijn Gielens a,*, Marnik G. Dekimpe b,1

a Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
b Catholic University Leuven, Naamsestraat 69, 3000 Leuven, Belgium

Received 3 March 2000; received in revised form 18 December 2000; accepted 23 March 2001

Abstract

The retailing industry, in the United States and Europe alike, faces maturing markets and stiffening domestic competition. In response, many of the industry's main players have shown a growing interest in cross-border initiatives. The success of such foreign entries obviously depends on the appropriateness of the retail firm's post-entry decisions, but may also depend on the strategic choices made at the time of entry, as they shape the platform from which competitive advantages can be gained. Little empirical evidence is available, however, on the contribution of these time-of-entry choices to post-entry performance, especially in the longer run.

In this paper, we consider simultaneously five strategic entry decisions: scale of entry, mode of entry, order of entry, the adaptation of the retail format to local market conditions and the familiarity of the store format to the parent company. We focus on the estimation of these decisions' effects on the long-run performance of a retail firm's foreign operations. In addition, we account for the moderating impact of three sets of control variables: characteristics of the foreign operations' retail mix, the parent firm's resources and the host-market's characteristics. Formally, the strategic choices and control variables are linked to the foreign operation's asymptotic performance level, which we derive from S-shaped growth models. We calibrate these models on a data set covering the post-entry performance of over 160 foreign entries made by Europe's top 75 food retailers, both towards Western-European countries and to a variety of transition economies in Eastern Europe.

The empirical findings suggest that the strategic decisions made at the time of entry continue to influence the foreign subsidiary's future performance, both in terms of sales performance and in terms of efficiency (sales/m²). Specifically, higher long-run post-entry sales and efficiency can be expected when entering early, with substantial scale, using no partners or acquired assets while offering a store format that is at the same time new to the host market and familiar to the parent firm. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Performance appraisal; International expansion; Retailing

* Corresponding author. Tel.: +31-13-466-2055; fax: +31-13-466-2875.
E-mail addresses: K.Gielens@kub.nl (K. Gielens), marnik.dekimpe@kuleuven.ac.be (M.G. Dekimpe).
1 Tel.: +32-16-326-957; fax: +32-16-326-732.
1. Introduction

Faced with maturing markets and stiffening competition, the retailing industry in both Europe and the United States has been forced to rethink and restructure itself in order to preserve its competitive position. In search of new opportunities, an important strategic option that can be taken is diversification, which can take place along two dimensions: across product boundaries and/or across market boundaries. The first dimension is reflected in the ongoing search for new store formats and/or new and broader store assortments. Still, to avoid a pure market-share game in increasingly saturated domestic markets, retailers are increasingly forced to also look for new geographical markets.

This may explain the recent growth in the number of cross-border initiatives by retail firms: the world’s 100 largest retailers are growing twice as fast abroad as domestically, and the 35 largest global retailers are each entering an average of one new market every year (Higgins, 1997; Mulhern, 1997). The Dutch supermarket chain Royal Ahold, for example, recently bought supermarkets in Poland (Eastern Europe), Spain (Western Europe) and both North and South (Argentina) America, while France’s Carrefour showed a similar internationalization urge with new operations in the Czech Republic, Chile and Indonesia, among others.

Notwithstanding this evolution, retailers are still struggling to develop the competencies needed to compete in the global arena (Kumar, 1997). Indeed, few international retailers realize comparable margins and/or returns abroad as in their home market, and few reach break-even volumes in their international operations. Carrefour loses money in many of its European and cross-continental markets, and also Wal-Mart realizes an average return on investment on its international operations that is substantially lower than in its American home market (The Economist, 1999). While the success of foreign entries clearly depends on the appropriateness of the firm’s post-entry decisions, the strategic choices made at the time of entry have been argued to also be of great importance, as they shape the platform from which competitive advantages can be gained, and hopefully sustained, over the subsidiary’s life cycle (Green et al., 1995).

Unfortunately, little empirical research is available on the relative contribution, if any, of different strategic choices at entry on post-entry performance, especially in the longer run (Feeer and Willard, 1990; Sharma and Kesner, 1996). This lack of empirical evidence may partly explain the considerable variability observed in the market place in terms of some key dimensions of the adopted foreign entry strategy. The Dutch Royal Ahold, for instance, primarily looks for the acquisition of, or joint ventures with, superior partners (interview with CEO C. Verhoeven, www.ahold.com), while the German Metro concern tends to enter through greenfield expansion. The British Tesco chain recently announced a change of strategy, in that it plans to grow less through acquisitions and more by opening its own greenfield sites (M + M Eurodata, 1993–2000). In terms of format choice, for many years France’s Carrefour only opened stores under its familiar hypermarket concept (M + M Eurodata, 1993–2000), while Aldi still operates exclusively through its hard-discount format. Both previous chains are therefore driven by supply-side considerations in their choice of format. Ahold and Tesco, in contrast, tend to adjust their store format to local market or demand conditions. Both realize most of their domestic sales through supermarkets, but have recently opened hypermarkets in Eastern Europe (Ahold, 1999, www.ahold.com; Tesco, 1999). As for the timing dimension, Tesco prefers to invest early in Central Europe to be able to steer market growth (Tesco, 1999), while Aldi only wants to enter after this market has sufficiently developed. As for entry scale, some retailers prefer to arrive in full force, as Carrefour did in Spain and Ahold on America’s East coast (The Economist, 1999), while Germany’s Rewe deliberately enters a market by taking over a small chain of stores from which it can expand later on (M + M Eurodata, 1993–2001).

Given this variability in entry strategy and the current ambiguity with respect to the long-run performance implications of the choices made at entry, we seek to answer in this paper questions such as: Is it better to quickly enter a market on a more limited scale, or to postpone entry until more resources have accumulated to enable a larger-scale commitment? What format should one adopt when entering a foreign market, the most popular one in the target...
country (demand consideration), or the one most familiar to the parent company (supply consideration)? Are foreign operations more successful when initiated through greenfield expansion, and/or does this depend on how quickly one enters the market, or on whether the entry is made in a transition economy?

Specifically, we consider simultaneously five strategic entry decisions: scale, mode and order of entry, the level of format adaptation to local market conditions and the extent of format familiarity on the part of the parent company. We focus on the estimation of these decisions’ main effects, as well as a variety of interaction effects, on the long-run performance of a retail firm’s foreign operations. In this estimation, we account for the impact of three sets of control variables: characteristics of the foreign operation’s retail mix, the parent firm’s resources, and characteristics of the selected host market. Formally, the strategic choices and control variables are linked to the foreign operation’s asymptotic performance level, which we derive from S-shaped growth models. We calibrate these models on a data set covering the post-entry performance of over 160 foreign entries made by Europe’s top 75 food retailers, both in Western-European countries and in a variety of transition economies in Eastern Europe.

As such, our research differs in three ways from previous studies on post-entry performance. First, we simultaneously consider multiple strategy components, while previous work has mostly concentrated on the impact of one or two of them. Second, attention is focused on the long-run performance consequences of these strategic choices. Finally, we focus on international entry decisions in the European retailing industry. In doing so, we expand the geographic coverage of the entry-strategy literature, which has mostly considered entry decisions into North-American markets, and answer Lieberman and Montgomery’s (1998) call for more research on the impact of entry strategies in service rather than consumer-goods industries.

The remainder of the paper is organized as follows. The conceptual framework and the ensuing hypotheses are developed in Section 2. Section 3 introduces the modeling technology, and Section 4 contains a discussion on our database. Section 5 presents our empirical findings that are validated in Section 6. Section 7 concludes with managerial implications of the findings and some areas for future research.

2. Hypothesis development

A basic premise of entry-strategy research is that the strategic decisions of the firm at entry continue to affect the post-entry performance of its foreign subsidiaries, even in the long run (Green et al., 1995). These decisions are not only hard to reverse; they also tend to shape the tactical options available to the company after entry (Sharma and Kesner, 1996). In Section 2.1., we discuss the hypothesized impact of five different entry decisions: (1) scale, (2) mode and (3) order of entry, along with the extent of (4) format adaptation and (5) format familiarity.

To properly evaluate the long-term impact of entry decisions on post-entry performance, we must take into account that the “strategic window” does not open at the same time for all players involved. Not all retail firms have the same resources to back their new ventures, and opportunities in different host markets may both differ and change over time. While the focus of our study is on the impact of different entry-strategy decisions, we control for several other factors to better assess the entry decisions’ contribution (see, e.g., Gatignon et al., 1990 for a similar conceptual approach). The controlling variables we consider in this study are introduced in Section 2.2.

2.1. Strategic entry decisions

Scale of entry. A variety of reasons may cause a positive relationship between entry scale and post-entry performance. First, large-scale entry may induce volume-driven cost advantages. The greater the scale of the initial set-up, the less growth is needed before the minimum-efficient scale of the industry is obtained (Biggadike, 1979). Second, large-scale entry (e.g., with many outlets) may ensure pre-emption of more attractive locations in both geographic and perceptual space (Lieberman and Montgomery, 1988). Third, incumbents are less likely to react aggressively when the entrant has made substantial hard-to-reverse investments, and other potential en-
Entrants may be less inclined to actually enter the market afterwards. The scale of the entry is therefore used as a signal of managerial commitment (Sharma and Kesner, 1996). Fourth, entrants holding more positive expectations are likely to make larger initial commitments (Caves, 1998). As such, entry scale may reflect the entrant’s performance expectations, and may already capture an option for further expansion. Finally, start-up scale can reflect the ability of the entrant to attract financial resources, a proxy for its future growth potential (Audretsch, 1995). Based on the above arguments, we hypothesize:

\[ H_1. \] Large-scale entry in an international market results in higher long-run performance in that market than smaller-scale entry.

**Mode of entry.** A firm seeking to operate abroad must determine the most appropriate mode, or institutional arrangement, for the new host-market entry (Anderson and Gatignon, 1986). Modes of entry differ in the degree of control the parent firm maintains over its foreign operations. Wilson (1980) and Delacroix (1993) find that entries through acquisition have a higher post-entry failure rate than entries through new ventures (i.e., greenfield expansion). This is typically attributed to (1) difficulties of integrating the acquired businesses into the parent system (see, e.g., Jemison and Sitkin, 1986), and (2) less managerial commitment to acquired outlets than to the ones management initiated (Li, 1995). As with acquisitions, entry through joint ventures involves the integration of at least two corporate cultures. Moreover, joint ventures may not only be troubled by differences in organizational culture, but also by difficulties in sharing property assets (Kogut and Singh, 1988). We therefore hypothesize:

\[ H_2. \] Entries in an international market through greenfield expansion result in higher long-run performance in that market than entries through acquisitions or joint ventures.

**Order of entry.** A substantial body of literature has documented the positive impact of (early) entry timing on subsequent performance (see, e.g., Kalyanaram et al., 1995). Most of the work in this area has been devoted to the introduction of new products in domestic markets, and less to how first-mover advantages translate into an international context (Lieberman and Montgomery, 1998). Some positive evidence is provided by Mascarenhas (1992a), who demonstrates that new products perform better in an international market if they are introduced early. A second “bias” in previous first-mover research is the focus on product introductions and the relative neglect of services (Lieberman and Montgomery, 1998). The latter is unfortunate, as managers of manufacturing firms have been found to view pioneering risks as more important than their counterparts in service industries, and as some causes of pioneering advantages, such as cost and differentiation advantages, may be less relevant in a service context. Still, in a recent survey, managers expect the pioneer to also have performance advantages in the services sector (Song et al., 1999), and we therefore hypothesize:

\[ H_3. \] Early entry in an international market results in higher long-run performance in that market than later entry.

The order of entry may not only have direct performance implications, there may also be an interaction effect with the two strategic decisions mentioned before: the scale and mode of entry. A substantial entry scale by early firms is thought to be effective in pre-empting the best locations, to

\[^2\] Some authors hypothesize a curvi-linear relationship between order of entry and post-entry performance, since early entrants may suffer from the liability of newness (see e.g. Golder and Tellis, 1993; Lilien and Yoon, 1990; Shankar et al., 1999; Zaheer, 1991), and could be leapfrogged by early followers. They argue that first entrants incur higher risks of entering the wrong country, may adopt a sub-optimal strategy in the right market, and could make costly mistakes in the process. Early followers, in contrast, may learn from the first movers’ mistakes and experience, while still avoiding a cluttered market with strong incumbents. In a retail setting, however, the preemption of the best (geographic) locations is of primary importance (Levy and Weitz, 1998), and we therefore hypothesize a monotonic relationship between order of entry and post-entry performance. We explicitly test for the absence of higher-order (non-monotonic) effects in the estimation section.

...
signal commitment for their move, and to create barriers for later entrants. Szymanski et al. (1995, p. 26) therefore hypothesize that the larger the capacity utilized by early movers, the greater the performance benefits of being early. In line with this argumentation, Nehrt (1996) finds empirical support for the notion that late entrants find it harder to digest large initial investments than innovators. Indeed, the former’s investments are likely to become less effective given the more competitive environment they encounter. We therefore argue that the positive impact of scale of entry is attenuated by the firm’s order of entry:

H4. The positive impact on performance of a large-scale entry into an international market is stronger for early than for late entrants.

With respect to the moderating effect of order on mode of entry, conflicting argumentation has been presented in the literature. On the one hand, being early offers an opportunity to select the best partners available in the target market, thereby mitigating the potential disadvantages of acquisitions and joint ventures (Szymanski et al., 1995). Moreover, greenfield expansion requires a larger and longer-term commitment, which may cause a reduction in flexibility. For example, joint-venture contracts can typically be renewed or renegotiated on shorter notice than the time needed to establish new entities in the foreign market (Pan et al., 1999). This flexibility is argued to be most critical in the early stages of market development (Szymanski et al., 1995). On the other hand, being early also offers the best opportunities to select the most suitable outlet locations, which is critical to the success of greenfield operations. As store location determines both the number and type of consumers a store will attract, being able to preempt the most attractive store sites may substantially increase the sales potential of the de novo stores (Levy and Weitz, 1998). This issue may be less important when dealing with acquisitions or joint ventures as one may still acquire or join forces with the earlier entrants that occupy the better locations. Given these equivocal arguments, we concur with Szymanski et al. (1995) that the sign of the interaction of order and mode of entry is best left as an empirical issue, and do not posit a directional hypothesis.

Format adaptation. The retail structure still differs greatly across national boundaries (Eurostat, 1998). Whereas the hard-discount format is widely accepted by German consumers (reflected in a market share exceeding 50%), it only represents a marginal share in markets such as the U.K. and France (a share of 5% to 10%). The hypermarket format is the most popular store concept in France, but is almost non-existing in the Netherlands. Given these differences, the question arises whether a retailer should adapt his store concept to the existing retail culture in the host market.

It has been argued that markets have become so homogenized that firms can market identical products and services (e.g., retail formats) around the globe through standardized marketing programs, and thereby capitalize on the company’s existing capabilities and knowledge (Hamel and Prahalad, 1985). Critics, however, argue that because of long-existing cultural, political and/or economic differences among nations, one should adapt marketing programs to the local market and environmental conditions (see, e.g., Parker and Tavassoli, 2000), implying that store formats which conform to local shopping and store-choice behavior will yield better post-entry performance. At present, little agreement exists whether and when standardization or adaptation is most appropriate in foreign markets. Because of these conflicting arguments, no directional hypothesis is formulated on the impact of format adaptation.

Format familiarity. The more remote the foreign operations are from the firm’s core product or business, the greater the uncertainty involved. In addition, the previously acquired expertise may not be directly transferable to the new setting (Caves, 1982; Feerer and Willard, 1990). While we restrict our attention to entries in the same line of business (grocery retailing), we still observe considerable variation in the choice of format. In the retailing industry, the strategic positioning is closely linked to the store format, such as hypermarket, supermarket or discount (Levy and Weitz, 1998). To reduce the uncertainty involved, and to benefit as much as possible from the parent firm’s experience, the parent firm may want to enter foreign markets with the
store format with which it is most familiar (Li, 1995). We therefore hypothesize:

**H5.** Entry in an international market with a store format more familiar to the parent company results in a higher long-run performance in that market than entry with a less familiar format.

As argued before, the probability that a retailer can successfully transfer its current physical and knowledge resources to a new market is higher if the business format used corresponds to the entrant’s core format. This positive effect of related entry is thought to be especially strong for greenfield operations. The more related the two formats, the more the firm is able to build on existing routines and the less likely it is that the firm will have to acquire externally the needed capabilities. Entry through acquisitions or joint ventures while using a familiar store format, in contrast, is likely to result in the purchase of unwanted or duplicate assets, and therefore result in less profitable outcome situations (Barkema and Vermeulen, 1998; Chatterjee, 1990).

**H6.** The positive effects on long-run performance of entering an international market with a store format familiar to the parent company are stronger in the case of greenfield expansion than in the case of entry through acquisitions or joint ventures.

Finally, the question arises whether the transfer of the retailer’s physical and knowledge resources is facilitated when the market is already used to that format. In that case, the firm’s marketing and logistic expertise is directly applicable without the need to first “educate” customers (and intermediaries) to not only switch stores but also store format (Kumar, 1997), making international scale economies easier to achieve (The Economist, 1999). This leads to the following hypothesis:

**H7.** The positive effects on long-run performance of entering an international market with a store format familiar to the parent company are stronger when that format is already well established in the target market.

### 2.2. Control variables

The long-term effectiveness of entry decisions is not tested in isolation; we correct for retail-mix, parent-firm, and environmental conditions. We present a brief motivation on their expected impact.

#### 2.2.1. Retail-mix characteristics

We control for three important dimensions of the retail mix: (1) the assortment size, (2) the pricing strategy adopted, and (3) whether certain additional services are offered.³

**Assortment size (+).** Increased demand for convenience raises consumers’ expectations to find a wide variety of items in their grocery store, i.e. to find what they want (Kahn and McAlister, 1997). Based on this market trend, we expect stores with a larger number of product lines to have a higher long-run performance than stores with a more limited assortment.

**Pricing strategy adopted: EDLP vs. HiLo (+ / −).** Retail pricing strategies are often labeled as EDLP (every day low prices on many categories) or HiLo (characterized by frequent, deep discounts on higher, regular prices—Shankar and Bolton, 1999). Retailers concerned about the lack of consumer credibility for “high–low” pricing are increasingly looking at an “every day low pricing” (EDLP) strategy to restore customer trust and loyalty. EDLP is also said to offer potential for improved customer service and merchandising, to result in better in-stock conditions, and to favor advertising that is more image rather than price oriented (Ortmeyer et al., 1991). HiLo strategies, in contrast, may increase store traffic, result in a more favorable price perception without the need to lower all prices, and increase sales on non-discounted items. Current empirical results are equivocal on the superiority of either technique, especially in the long run (see Kahn and McAlister, 1997 for a recent review).

**Additional services: loyalty cards (+).** Some consumers are not particularly motivated by price considerations, but are more sensitive to service-based strategies. One such strategy is the use of

³ The sign of the expected impact is presented between parentheses.
loyalty cards to bond customers to the company, its products and/or services. Not only do these programs attempt to maintain current sales, they also try to increase the lifetime value of the existing customer base by increasing its loyalty and by stimulating cross-selling (Dowling and Uncles, 1997).

2.2.2. Parent-firm resources and skills

The resource bases taken into consideration in this study are (1) the retailer’s size, (2) the scope of his international experience, and (3) his assortment policy, as reflected in the share of his private labels.

**Firm size (+).** Entrants are expected to perform better when their parent firm possesses the skills and resources critical for competitiveness vis-à-vis incumbents and other potential new entrants (Sharma and Kesner, 1996). As larger firms tend to have more market power, and as they enjoy economies of scale and scope in their supplier management and information collection/processing (Mascarenhas, 1992b; Tan and Vertinsky, 1996), they can be expected to perform better.

**International experience (+ / −).** Firms that already have international experience may be able to capitalize on these experiences when entering other markets (Johansson and Vahlne, 1977). Previous work suggests that organizational learning through experience may reduce the risks involved in new entries, and benefit subsequent market performance and survival probability. Alternatively, one could argue that firms operating in many markets have numerous expansion possibilities, and will have to spread their limited (managerial and/or financial) resources across a broader set of options (Sharma and Kesner, 1996). This spreading of resources may negatively impact the performance of individual foreign entries.

**Assortment policy; private label share (−).** An important evolution in the retailing industry is the growth of store brands. Investments in the development of a private label offering not only reflect a commitment to the local (domestic) market, but may also limit the resources available for foreign expansion (Sharma and Kesner, 1996). In addition, when they want to carry manufacturer-branded products in the foreign markets, retailers with a large private-label share often have a harder time in the negotiations with their suppliers than other retailers (Kumar, 1997). Finally, from the consumer’s point of view, retailers who rely heavily on private labels not only have to convince the host population to switch stores, but to also switch brands (Kumar, 1997). These retailers are therefore expected to perform worse in new host markets.

2.2.3. Host market characteristics

The intrinsic attractiveness of the host market is captured in five ways: (1) the size of its population, (2) the purchasing power or wealth of its inhabitants, (3) its competitive structure, (4) the transitional nature of its economy, and (5) its cultural and geographic distance to the home country and/or other countries the parent firm is operating in already.

**Population (+).** Market size is often used as the economic rationale of foreign market entry. Not only are retail sales closely linked to the number of people living in a given servicing area, the average cost of serving larger markets will typically be smaller and decrease faster than when servicing smaller markets (Davidson, 1980).

**Wealth (+).** Post-entry performance is expected to be related to a country’s wealth and stage of economic development. Specifically, international expansion in the retailing sector has been argued to be closely linked to the evolution (in size and income level) of the middle class in the host market (Higgins, 1997).

**Competitive structure: number of competitors (−) and variation in market shares (+).** It has been argued that intense rivalry in the host market not only serves as a barrier to entry, but also affects the post-entry performance of firms that try to overcome this barrier. In general, retailers face different levels of competition in their new host market depending on (a) the number of incumbents they encounter, and (b) the heterogeneity in the latter’s market shares (Dhar and Hoch, 1997); the larger the number of firms and the more equal their size, the higher the

---

4 Even though economies of scale could arise at the production and logistics side in case of a widespread use of the private label across multiple countries, we follow the common notion that the demand-side disadvantages may outweigh these supply-side benefits.
likelihood of intense rivalry. As such, the most favorable environment would be one in which there are relatively few retailers, with one or a few of them dominant. In such markets, competition in both the purchase and sales market is expected to be less intense and more orderly (Galbraith and Stiles, 1983), which should benefit post-entry performance.

Transitional nature of the economy (+/−). Emerging markets are said to offer a major growth opportunity (Arnold and Quelch, 1998). The rapid pace of economic development, along with government policies favoring economic liberalization, offers great perspectives to foreign investors. However, at present, macro-economic and political stability have not yet been achieved in most transition economies. Together with missing institutional features, this instability increases uncertainty and risks for foreign investors (Hoskisson et al., 2000). Because of this ongoing uncertainty, it is hard to predict whether transition economies will ever live up to their expectations.

Cultural and physical distance (−). A greater distance between the host and home market leads to more costs of coordination and less benefits of scale. As for cultural distance, it is believed that the routines and repertoires that generate a firm’s competitive advantage appear to be embedded and constrained in national culture, which make them difficult to replicate in other cultures (Hofstede et al., 1990). Physical proximity, on the other hand, lowers costs of managerial co-operation and may result in greater benefits from economies of scale, spillover effects and standardization of marketing activities. Moreover, the international diffusion literature stresses the ability of cross-country consumer learning. Cross-border shopping and the ability of consumers in one country to communicate with potential consumers in the second market provide an additional source of information, which may influence the take-off and the long-run success of the new retail subsidiary (Gatignon et al., 1989).

3. Methodology

Previous studies have measured post-entry performance at different points in time, varying from a low of 1 to 2 years after entry (Luo, 1998) to a high of over 10 years (Mascarenhas, 1992a), depending on the type of product/market entry under investigation. Both approaches result in a loss of useful information. The former only uses performance information shortly after the initial entry, and ignores the additional information on entries that took place a longer time ago. This information is particularly relevant given our interest in the long-run performance consequences of the entry decisions. Only using information on entries undertaken a long time ago neglects all information on the more recent entries, and causes a reduction in sample size. Some authors have addressed the issue by using the last available performance observation for every entry (e.g., Mascarenhas, 1992b). This approach, however, is likely to compromise the comparability of performance estimates across entries of different length, as one might deal with entries in a different stage of their life cycle.

In this study, we define long-run post-entry performance as the asymptotic performance level the foreign operation is converging to, which we quantify through the saturation level of an S-shaped growth curve fitted to the entries’ over-time performance data. In estimating these growth curves, we incorporate all available data points, and by focusing on the estimated asymptotic performance level, the measure becomes time independent. In this way, we ensure the comparability of the estimates across entries with different starting dates. Growth curves have been used to model and forecast the evolution of a variety of phenomena, including the development of new markets (Meade and Islam, 1998) and the evolution of new product sales (Franses, 1994a). They typically involve three to four parameter models to write the evolution of a performance measure as a deterministic function of time. One of the more popular and better performing (Meade and Islam, 1998) models is the Gompertz curve. For the performance evolution of a given entry, the Gompertz

5 Conceptually, our approach is related to, e.g., Dekimpe et al. (1998), Parker and Gatignon (1994, 1996) and Shankar et al. (1998), who relate the market-potential parameter in diffusion models to a set of predictor variables.
log \( P_{ijt} \) = log \( P_{ijt-1} \) = \( \alpha [\log M_{ij} - \log P_{ijt-1}] + e_{ijt} \),
\( M_{ij} \) is the key construct of interest, i.e. the asymptotic, relative to the entry date of an entry by firm \( i \) when entering market \( j \); we pool the observations across all entries, and the errors of the observations belonging to one firm and by host market. In such a pooled data set, the errors of the observations belonging to one firm and/or market may still be correlated due to omitted variables, and a pooling bias may emerge. To account for this stratification, a fixed-effects correction is used which results in the further addition of firm (FD) and host-country (CD) dummy variables to Eq. (2) (see, e.g., Gatignon et al. 1990 for an in-depth discussion). We hereby assume that each entry-specific error can be additively decomposed into a company- and country-specific component.

Third, apart from these time-invariant control variables, we also want to capture the overall evolution in the grocery retail market (IND) in a given country. By including a time-varying demand-shift variable (see Kadiyali et al., 2000 for a similar practice), we control for economy (industry) wide shocks whose effects should not be confounded with the effects of the initial strategic choices. In doing so, the asymptotic performance level is allowed to change over time, and denoted as \( M_{ijt} \).

Upon incorporation of the above three extensions, Eq. (2) becomes:

\[
\log M_{ijt} = \beta_0 + \sum_{i=2}^{I} \gamma_i FD_i + \sum_{j=2}^{J} \delta_j CD_j + \beta_1 \log(Scale_{ij}) + \beta_2 \log(Mode_{ij})
\]

Eq. (3) is non-linear in the parameters, and therefore calibrated by non-linear estimation techniques.

Eqs. (2) and (3) are subsequently augmented in three ways. First, we want to control for retail-mix, i.e. assortment size (ASIZE), the presence of a loyalty card (LOYCARD) and the use of an every-day-low-price strategy (EDLP), parent-firm, i.e. its size (FSIZE), international experience (EXP) and extent of private-label use (PRIVLAB), and target-market, i.e. population (POP), wealth (WEALTH), number of competitors (COMP), market-share variance (MSVAR), the transitional nature of the economy (TRANS) and the cultural (CULT) and physical (PHYS) distance. Each of these characteristics is added to Eq. (2) as a time-invariant covariate (typically measured in the year prior to the entry; see Section 4).

Second, our data set covers international entries by approximately 40 retail firms in over 20 target markets, causing our observations to be stratified by firm and by host market. In such a pooled data set, the errors of the observations belonging to one firm and/or market may still be correlated due to omitted variables, and a pooling bias may emerge. To account for this stratification, a fixed-effects correction is used which results in the further addition of firm (FD) and host-country (CD) dummy variables to Eq. (2) (see, e.g., Gatignon et al. 1990 for an in-depth discussion). We hereby assume that each entry-specific error can be additively decomposed into a company- and country-specific component.

Upon substitution of (2) into (1), the following estimation equation is obtained:

\[
\log M_{ij} = \beta_0 + \beta_1 \log(Scale_{ij}) + \beta_2 \log(Mode_{ij})
\]

Upon substitution of (2) into (1), the following estimation equation is obtained:

\[
\log M_{ij} = \beta_0 + \beta_1 \log(Scale_{ij}) + \beta_2 \log(Mode_{ij})
\]

Upon substitution of (2) into (1), the following estimation equation is obtained:

\[
\log M_{ij} = \beta_0 + \beta_1 \log(Scale_{ij}) + \beta_2 \log(Mode_{ij})
\]
+ $\beta_3 \log(\text{ORDER}_{ij})$
+ $\beta_4 \log(\text{ORDER}_{ij}) \cdot \log(\text{SCALE}_{ij})$
+ $\beta_5 \log(\text{MODE}_{ij})$
+ $\beta_6 \log(\text{ADAPT}_{ij}) + \beta_7 \log(FAMI_{ij})$
+ $\beta_8 \log(\text{FAMI}_{ij})$
+ $\beta_9 \log(\text{FAMI}_{ij}) \cdot \log(FAMI_{ij})$
+ $\beta_{10} \log(\text{ASIZE}_{ij}) + \beta_{11} \log(\text{LOYCARD}_{ij})$
+ $\beta_{12} \log(FSIZEx) + \beta_{13} \log(\text{FSIZE}_{ij})$
+ $\beta_{14} \log(\text{ORDER}_{ij}) + \beta_{15} \log(\text{PRIVLAB}_{ij})$
+ $\beta_{16} \log(\text{POP}_{ij}) + \beta_{17} \log(\text{WEALTH}_{ij})$
+ $\beta_{18} \log(\text{COMP}_{ij}) + \beta_{19} \log(\text{MSVAR}_{ij})$
+ $\beta_{20} \log(\text{TRANS}_{ij}) + \beta_{21} \log(\text{CULT}_{ij})$
+ $\beta_{22} \log(\text{IND}_{ij})$

which is then substituted into (1) as before.

To avoid over-parameterization and/or multicollinearity problems with e.g. the fixed transition-economy dummy variable, we propose not to include the full set of fixed-effects dummy variables. Instead, an iterative bottom-up procedure (i.e., starting one dummy at a time, and gradually adding additional dummy variables) is used to identify the significant ones.

4. Data description and measurement

4.1. Sample composition

Information on post-entry performance was available for all entries made from 1988 onwards by the top 75 European grocery retail firms towards other Western and Eastern European markets. In what follows, we briefly describe and motivate the different dimensions of the data set: (1) the sample of firms, (2) the scope of host or target markets, and (3) the observation window.

Foreign entries made by the top players in the grocery retailing industry are considered, provided that these involved the firms’ food operations. The latter restriction is imposed to obtain a more homogeneous sample; it eliminates, e.g., the comparison of foreign market entries in the computer business (e.g., El Cortes in the Czech Market) with entries in the do-it-yourself business (e.g., Metro in Greece). Within the food retailing sector, 45 of Europe’s top 75 players, coming from 12 different home markets, made one or more international entries over the considered time span.

Over 160 of these entries were directed toward other (Western and Eastern) European countries. We restrict ourselves to entries into European markets as cross-continental moves are still quite rare, often restricted to a few firms, and either very recent or scarcely documented. Still, the sample incorporates international entries into 24 different host countries that cover both Western and Eastern European markets.

Finally, we consider all international entries made from the late 1980s until the end of 1998. It is generally accepted (see, e.g., Barth et al., 1996) that the grocery retailing industry has experienced two internationalization waves. A first wave, situated in the 1970s and early 1980s, consisted primarily of expansion into adjoining countries and typically involved equity investment or acquisition. The second wave, starting in the late 1980s and still accelerating, also involves movements beyond a retailer’s established trading area, and also comprises greenfield expansion and joint ventures. It is the latter, still ongoing, wave that is studied.

Information on the internationalization decisions of the different firms was obtained from $M+M$ Eurodata, and was extensively cross-validated through searches of Reuters, company web sites, annual reports, and local trade publications (e.g., Distributie Vandaag for Belgium or Points de Ventes in France). All international entries through green-
Table 1
Overview retail entries included in study

<table>
<thead>
<tr>
<th>Home region</th>
<th>Benelux</th>
<th>Central Europe</th>
<th>France</th>
<th>Germany, Austria and Switzerland</th>
<th>Russia and Baltic States</th>
<th>Scandinavia</th>
<th>Southern Europe</th>
<th>Southeast Europe</th>
<th>UK and Ireland</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benelux, e.g. Ahold (NL), Delhaize (B), GIB (B)</td>
<td>1</td>
<td>18</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>France, e.g. Carrefour, Intermarché, Casino</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>Germany, Austria and Switzerland, e.g. Metro (D), Migros (CH), Spar (A)</td>
<td>3</td>
<td>33</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>70</td>
</tr>
<tr>
<td>Scandinavia, e.g. Kesko (FIN), Reitan (N), Dansk Supermarked (DK)</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Southern Europe, e.g. JMR (P)</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>UK and Ireland, e.g. Tesco, Booker, Costcutter</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7</td>
<td>77</td>
<td>11</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>43</td>
<td>7</td>
<td>6</td>
<td>169</td>
</tr>
</tbody>
</table>

Note: The Benelux countries are Belgium, The Netherlands and Luxembourg. The Scandinavian countries include Denmark, Finland, Norway and Sweden. The Southern European countries investigated in this study are Greece, Italy, Spain and Portugal. The Central European countries are The Czech Republic, Hungary, Poland and Slovakia. The Southeastern markets under investigation are Bulgaria, Romania and Slovenia.

Field expansion and acquisitions were considered, while joint ventures were included when a participation of more than 50% is taken.

Table 1 presents summary information on the international entries in our sample. To keep the table manageable, we grouped host and home markets into meaningful broader categories; for our subsequent empirical analyses, however, both the dependent and independent measures are considered at the country level (see below for details).

4.2. Variable operationalization

4.2.1. Post-entry performance

In this study, two performance indicators, (1) an output and (2) an efficiency measure, are used to evaluate a retail subsidiary’s post-entry performance. The output dimension is captured through the total annual sales series of the foreign subsidiary, obtained
by adding the sales of all outlets the chain has in that country (all sales are expressed in Euro, and deflated to a common base year). Efficiency is typically operationalized as an output-to-input ratio, where the inputs represent the investments into the foreign operations. We operationalize the latter through the total store surface of the chain in a given country, resulting in a sales per m$^2$ efficiency measure.

Within marketing, output measures such as sales (Green and Ryans, 1990; Shankar et al., 1998) and market share (Kalyanaram et al., 1995) are most often used to assess post-entry performance. We focus on the former, but control for the overall evolution in a country’s retail industry by incorporating the total retail grocery sales in a given country (minus the sales of the considered entry) as a time-varying demand-shift control variable (see Kadiyali et al., 2000 for a similar practice).

In a retail context, a number of physical output measures have been proposed to quantify the performance of established chains/outlets, such as the number of transactions performed, the number of items sold, the value added, and sales (see, e.g., Bucklin, 1978). Ingene (1982) discusses the pros and cons of each of these output measures. He concludes that the usefulness of each measure depends on the researchers’ objectives, but that, in general, sales can be considered a ‘good measure’ of retail performance (p. 77).

To assess a firm’s economic performance, efficiency measures are often thought to be a valuable alternative/complement. While intuitively appealing, output-to-input ratios have been used only rarely thus far to evaluate post-entry performance (see Anderson and Gatignon, 1986 for a rare, conceptual, exception). In the retailing literature, however, efficiency has been a standard way to evaluate the performance of established outlets and chains. A critical issue in this respect is the selection of the relevant input series. In this study, we use the stores’ surface, aggregated across all outlets of the chain in a given country, as input measure. As both dependent variables, sales and efficiency, may be driven by some common random variation (and since the lagged dependent variable differs in both equations), non-linear SUR is used to exploit potential efficiency gains from their joint estimation.

Given our interest in estimating long-term, asymptotic performance levels from aggregate growth models, the question arises how many data points are needed for reliable estimates. In the context of the Bass model, Schmittlein and Mahajan (1982) suggest to either have more than 10 data points per entry or observations past the inflection point. As illustrated in Fig. 1, the inflection point of the two performance measures lies around the fourth year after entry, and many observations past this inflection point are available.

4.2.2. Strategic entry decisions

Scale of entry. To measure the scale of entry, we use the store surface at entry. The sample average is 12,985 m$^2$, with a range from 300 to 323,531 m$^2$.

Mode of entry. To capture the mode of entry decision, a dummy variable is created to indicate whether or not a retailer entered through greenfield operations (67% of all entries). This figure confirms that the second internationalization wave in the retailing industry is indeed mainly characterized by greenfield operations, as pointed out by Barth et al. (1996).

Order of entry. We define entry timing as the order in which a retailer entered a foreign market relative to all other foreign players, i.e. first and

---

10 Such an aggregation is common practice when comparing performance across foreign entries (see, e.g., Hultman and McGee, 1989; Tan and Vertinsky, 1990).
11 Alternatively, one could envision working with a relative performance measure (e.g., market share rather than sales) as dependent variable. In unreported analysis, we obtained similar substantive results (apart from the order of entry×mode of entry interaction which became insignificant) when working with relative (market share) rather than absolute (sales) performance levels.
12 Similar graphs are obtained when looking at the performance evolution of most individual entries (rather than at the evolution of the mean performance), and are available from the authors upon request. We should point out, however, that for some of the shorter series, no data past the inflection point were available (even though many were in the pooled data set, as shown in Fig. 1). We will, however, extensively validate the sensitivity of our substantive findings to these shorter series in Section 6.1.
Fig. 1. Sales and efficiency evolution (means). Note: The numbers of observations on which the mean is computed at each different year after entry \((t = 1−8)\) amount, respectively, to 161, 139, 126, 99, 77, 57, 41, and 26. The total number of observations in the data set is 753. For some of the earliest entries, more than eight observations were available. When plotting their individual performance graphs, a similar S-shaped pattern was obtained.

second wave entrants, in that market.\(^{13}\) There is high variability in this entry-decision variable. The mean is approximately 6, i.e. on average, retailers expanding abroad not only have to compete with the local incumbents, but also with five other international players. Some entrants, however, did not yet have to deal with other international players at the time of their entry \((\text{min } = 1)\), while others had to cope with up to 25 international retail investors \((\text{max } = 26)\), as was the case for the late entrants in Poland.

Format adaptation. The extent of product adaptation is measured through the market share in the target market \((\text{in the year preceding the entry})\) of the format used to enter that market. Four different formats are considered in this respect: hard discount\(^{14}\), supermarket, hypermarket and cash and

\(^{13}\) VanderWerf and Mahon (1997) report in their meta-analysis on first-mover advantages that most studies use a similar operationalization \((\text{i.e. one for the first entrant, second for the second, etc.})\).

\(^{14}\) Hard discounters are food stores working to the discount principle, i.e. limited product range, basic store fittings and low prices.
carry. A value of 1 for the adaptation variable would express that entry occurs with a format that completely dominates the host country’s grocery market. A value of 0, on the other hand, implies that entry occurs with a format that is completely new to the market. On average, entry occurs with a format which represents 26.3% of the grocery market. In some cases, however, entry occurs with a format that completely dominates the host market (min = 0), while in other markets one chooses to enter with a format that already has a strong presence (max = 73%).

**Format familiarity.** The parent firm’s familiarity with the selected format is measured through the percentage of worldwide sales the parent firm realizes (again in the year preceding entry) through that format. For example, a value of 50% indicates that half of the parent firm’s sales are obtained from operations with the format used in entering the new host market. For some retailers, the format choice represents a true diversification as they have no previous experience with the format (min = 0) while others are clear mono-format retailers (max = 1). On average, entry occurs with a store concept that represents 55% of current sales.

### 4.2.3. Control variables

**Retail mix.** To control for differences in the retail mix, measures for assortment size, pricing policy and service orientation were included. Information on these covariates was again obtained from M+M Eurodata, and validated (augmented) through annual reports and/or trade publications. Assortment size was operationalized through the number of product lines available in a typical store of the foreign retail venture. The nature of the subsidiaries’ pricing policy was captured through a dummy variable for a (predominant) use of an every-day-low-pricing policy, while another dummy variable was used to indicate the presence of loyalty cards in the subsidiary’s service offering. Because of data limitations, these covariates were treated as time-invariant.

**Parent firm skills and resources.** To assess the influence of the parent firm on the long-term performance of its foreign subsidiary, three indicators are included: (1) the sales of the parent firm to measure its size, (2) the number of countries a retailer is operating in world-wide as a proxy for the scope of its international experience\(^{15}\), and (3) the share of private labels in the parent retailer’s home assortment. All variables are measured at the time of entry.

**Host-market attractiveness.** To measure host-market attractiveness, the following time-invariant (computed at the time of entry) indicators were used: (1) GNP per capita, (2) the number of inhabitants in the host market, (3) the competitive intensity in the market, as reflected in the number of players in the market with a market share exceeding 1% and the market-share variation across those players (see Dhar and Hoch, 1997 for a similar conceptualization), (4) a dummy variable to identify entries into Eastern Europe, and (5) two distance measures: an indicator variable for previous activities in a neighbouring country (reflecting physical distance) and a measure of cultural distance between home and target country.\(^{16}\) The required information for GNP/Capita and population size was obtained from the Worldbank Statistics (Worldbank, 1990–1998), and the data to construct the competitiveness and familiarity indices were obtained from M+M Eurodata.

Table 2 summarizes the descriptive statistics for both the strategic entry variables and all control variables. Apart from the discrete mode-of-entry variable, all strategic entry decisions were mean-centered\(^{17}\) to reduce excessive correlations with the interaction terms (see, e.g., Jaccard et al., 1990). All correlations between the independent variables were then well below 0.8, while the variance-inflation

\(^{15}\) The number of countries one is operating in is thought to be a better indicator of the breadth of the firm’s international operations than the amount of foreign sales, as the latter can come disproportionately from one country (Tallman and Li, 1996).

\(^{16}\) Cultural distance is derived form Schwartz’ national-culture scores for Western (Schwartz and Ros, 1995) and Eastern European (Schwartz and Bardi, 1997) countries. A distance measure (Kogut and Singh, 1988) is then used to capture the degree to which the cultural norms in one country differ from those in another country (details are available from the authors upon request).

\(^{17}\) The strategic entry parameters therefore measure the impact from deviations from the mean, and the main effect parameters measure the total impact of a change in that variable assuming the interacting variable stays at the mean value.
Table 2
Descriptive statistics

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard deviation</th>
<th>Range</th>
</tr>
</thead>
</table>

**Independent variables**

- **Entry decisions**
  - Scale of entry
    - Mean: 12,985
    - Standard deviation: 28,952
    - Range: 300–323,531
  - Mode of entry: Greenfield
    - Mean: 66.8%
    - Standard deviation: –
    - Range: –
  - Order of entry
    - Mean: 6.3
    - Standard deviation: 5.3
    - Range: 1–26
  - Format adaptation
    - Mean: 26.3%
    - Standard deviation: 17.0%
    - Range: 0.0%–73.0%
  - Format familiarity
    - Mean: 55.0%
    - Standard deviation: 34.0%
    - Range: 0.0%–100%

- **Controlling variables**
  - Retail mix
    - Assortment size
      - Mean: 14,950
      - Standard deviation: 20,871
      - Range: 540–100,000
    - Pricing strategy: EDLP
      - Mean: 23.7%
      - Standard deviation: –
      - Range: –
    - Loyalty card
      - Mean: 23.1%
      - Standard deviation: –
      - Range: –
  - Parent firm resources and skills
    - Firm size (million Euro)
      - Mean: 11,627.0
      - Standard deviation: 10,221.0
      - Range: 2124–54,594
    - International experience
      - Mean: 3.9
      - Standard deviation: 3.9
      - Range: 0–24
    - Private label share
      - Mean: 26.0%
      - Standard deviation: 22.0%
      - Range: 1.0%–94.0%
  - Host market characteristics
    - Wealth: GNP/capita
      - Mean: 10,996.0
      - Standard deviation: 9624.4
      - Range: 1250–48,440
    - Population million inhab.
      - Mean: 28.9
      - Standard deviation: 26.7
      - Range: 0.4–148.9
    - Competitive structure
      - Variation market share
        - Mean: 0.03
        - Standard deviation: 0.03
        - Range: 0.0–0.15
      - Number of competitors
        - Mean: 7.8
        - Standard deviation: 5.2
        - Range: 0.0–17.0
      - Transitional nature economy
        - Mean: 51.0%
        - Standard deviation: –
        - Range: –
    - Distance
      - Cultural distance
        - Mean: 1.2
        - Standard deviation: 0.4
        - Range: 0.0–1.9
      - Physical proximity
        - Mean: 64.0%
        - Standard deviation: –
        - Range: –
  - Demand shifter (million Euro)
    - Mean: 28,998
    - Standard deviation: 40,801
    - Range: 67–181,458

**Dependent variables**

- Sales (million Euro) (overall)
  - Mean: 179.0
  - Standard deviation: 27.6
  - Range: 0.5–1800
- Efficiency = sales per m²
  - Mean: 0.005
  - Standard deviation: 0.004
  - Range: 0.0002–0.06

*a* 0–1 variables. We report the proportion of the observations having the value 1.

5. Empirical findings

5.1. Sales performance

In terms of long-run sales performance, we observe that our hypotheses regarding the main effects of the strategic entry decisions are corroborated (Table 3). Significant long-run impacts are found for all five entry decisions: a positive effect is found for size of entry (*p* < 0.01), greenfield operations perform better than acquisitions and joint ventures (*p* < 0.01), a monotonically declining relationship is ob-

---

As indicated before, only the significant fixed-effect dummy variables were kept in the model. Specifically, eight firm dummies and 10 host-market dummies turned out to be significant in the sales regression. In the efficiency equation, 12 firm dummies and five host-market dummies were significant.
Table 3
Estimation results: Gompertz model

<table>
<thead>
<tr>
<th>Expected effect</th>
<th>Sales</th>
<th>Efficiency</th>
<th>β</th>
<th>t-value</th>
<th>β</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategic entry decisions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale of entry (β₁)</td>
<td>+</td>
<td>0.35</td>
<td>5.92</td>
<td>0.08</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td>Mode of entry (β₂)</td>
<td>+</td>
<td>0.96</td>
<td>5.89</td>
<td>0.41</td>
<td>4.01</td>
<td></td>
</tr>
<tr>
<td>Order of entry (β₃)</td>
<td>–</td>
<td>–0.39</td>
<td>–3.73</td>
<td>–0.24</td>
<td>–3.59</td>
<td></td>
</tr>
<tr>
<td>(Order of entry) × (Scale of entry) (β₄)</td>
<td>–</td>
<td>–0.04</td>
<td>–0.73</td>
<td>–0.04</td>
<td>–0.97</td>
<td></td>
</tr>
<tr>
<td>(Order of entry) × (Mode of entry) (β₅)</td>
<td>+/–</td>
<td>–0.31</td>
<td>–1.95</td>
<td>–0.10</td>
<td>–1.00</td>
<td></td>
</tr>
<tr>
<td>Format adaptation (β₆)</td>
<td>+/–</td>
<td>–0.26</td>
<td>–4.83</td>
<td>–0.07</td>
<td>–2.19</td>
<td></td>
</tr>
<tr>
<td>Format familiarity (β₇)</td>
<td>+</td>
<td>0.10</td>
<td>2.51</td>
<td>0.05</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>(Format familiarity) × (Mode of entry) (β₈)</td>
<td>+</td>
<td>0.17</td>
<td>2.32</td>
<td>–0.01</td>
<td>–0.27</td>
<td></td>
</tr>
<tr>
<td>(Format familiarity) × (Format adaptation) (β₉)</td>
<td>+</td>
<td>0.04</td>
<td>1.72</td>
<td>0.02</td>
<td>1.44</td>
<td></td>
</tr>
<tr>
<td><strong>Retail mix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assortment size (β₁₀)</td>
<td>+</td>
<td>0.04</td>
<td>0.75</td>
<td>0.18</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td>Loyalty card (β₁₁)</td>
<td>+</td>
<td>0.82</td>
<td>4.48</td>
<td>0.31</td>
<td>2.47</td>
<td></td>
</tr>
<tr>
<td>Pricing strategy (β₁₂)</td>
<td>+/–</td>
<td>0.49</td>
<td>2.66</td>
<td>0.15</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td><strong>Parent-firm resources and skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm size (β₁₃)</td>
<td>+</td>
<td>0.11</td>
<td>1.44</td>
<td>0.21</td>
<td>3.68</td>
<td></td>
</tr>
<tr>
<td>International experience (β₁₄)</td>
<td>+/–</td>
<td>0.09</td>
<td>2.32</td>
<td>0.07</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td>Private label share (β₁₅)</td>
<td>–</td>
<td>0.04</td>
<td>0.78</td>
<td>–0.13</td>
<td>–1.98</td>
<td></td>
</tr>
<tr>
<td><strong>Host-country characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population (β₁₆)</td>
<td>+</td>
<td>0.97</td>
<td>4.05</td>
<td>0.57</td>
<td>4.53</td>
<td></td>
</tr>
<tr>
<td>Wealth (β₁₇)</td>
<td>+</td>
<td>1.26</td>
<td>2.32</td>
<td>0.19</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Competitive structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of competitors (β₁₈)</td>
<td>–</td>
<td>0.02</td>
<td>0.27</td>
<td>–0.09</td>
<td>–2.47</td>
<td></td>
</tr>
<tr>
<td>Variation market shares (β₁₉)</td>
<td>+</td>
<td>0.22</td>
<td>1.14</td>
<td>0.22</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>Transitional nature economy (β₂₀)</td>
<td>+/–</td>
<td>–1.34</td>
<td>–2.06</td>
<td>–0.79</td>
<td>–2.69</td>
<td></td>
</tr>
<tr>
<td><strong>Distance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cultural distance (β₂₁)</td>
<td>–</td>
<td>0.76</td>
<td>1.19</td>
<td>0.50</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Physical proximity (β₂₂)</td>
<td>+</td>
<td>0.01</td>
<td>0.08</td>
<td>0.11</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td><strong>Demand shifter (β₂₃)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.38</td>
<td>1.09</td>
<td>0.38</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.01 (one-sided).
*p < 0.05 (one-sided).
*p < 0.1 (one-sided).
*p < 0.01 (two-sided).
*p < 0.05 (two-sided).
*p < 0.10 (two-sided).

served with order of entry (p < 0.01)\(^{19}\), it is better to offer a new format to the host market (p < 0.01), and entries with store formats familiar to the parent company also fare better (p < 0.01). As such, higher long-run post-entry sales are expected when entering early, with substantial scale, using no partners or acquired assets while offering a store format that is at the same time new to the host market but familiar to the parent firm.\(^{20}\)

\(^{19}\) We extensively tested for higher-order (quadratic) effects, but the quadratic term was never close to significant, neither in the main effect nor in the interaction effects with scale or mode of entry. Details are available from the authors upon request.

\(^{20}\) It is interesting to note that these substantive findings were unaffected when additional control dummy variables for the subsidiaries’ store format (e.g., hypermarket, discount) were included. This was also the case for the efficiency results.
in the choice-of-format decision, we see that the former are relatively less important in case of joint ventures and acquisitions (elasticity of 0.13, versus −0.23 for format adaptation). An explanation for this finding is that in this entry mode, the expertise of the parent company may be less relevant than that of the local partners. When dealing with greenfield expansion, in contrast, the expertise and knowledge of the parent firm tends to be much more important. This is reflected in an elasticity of 0.29, which is of comparable order of magnitude as the format-adaptation decision.

With respect to the control variables, we find that two retail-mix variables have a positive significant effect on long-run sales performance, i.e. (1) the use of loyalty cards, \((p < 0.01)\) and (2) the adopted pricing strategy, where EDLP outperforms HiLo \((p < 0.01)\). The latter finding extends the experimental findings of Drèze et al. (1994) that EDLP increases (short-run) volume. As for parent-firm skills and resources, we first observe that substantial international experience at the time of entry positively impacts the future sales evolution of the firm’s foreign subsidiaries \((p < 0.05)\). Second, firm size is marginally significant \((p < 0.1)\), indicating that backing by a large-sized parent firm helps the new subsidiary, even in the long run. Three of the host-market control variables have a significant impact: (1) the size of the host-market population \((p < 0.01)\), (2) the wealth of the host market as measured by its GNP/capita \((p < 0.05)\), and (3) the transitional nature of the market, where we find that the risks attached to emerging economies may have lasting negative effects in the long run \((p < 0.05)\).

The significant negative effect for this last dummy variable indicates that, all else equal, managers should anticipate a lower long-run performance level when entering into Eastern European countries. The dummy variable only allows for an intercept differ-

Table 4
Elasticities evaluated at the mean

<table>
<thead>
<tr>
<th></th>
<th>Sales</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenfield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operations</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Joint ventures and acquisitions</td>
<td>-0.35</td>
<td>-0.24</td>
</tr>
<tr>
<td>Order of entry</td>
<td>-0.70</td>
<td>-0.39</td>
</tr>
<tr>
<td>Format adaptation</td>
<td>-0.23</td>
<td>-0.23</td>
</tr>
<tr>
<td>Format familiarity</td>
<td>0.29</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Apart from these main effects, we also found support for three interaction effects. Order of entry attenuates the positive impact of greenfield operations \((p < 0.10)\), while the resource and knowledge transfer that arises when entering with a familiar store format is most pronounced in case of greenfield expansion \((p < 0.05)\) and when the market is already used to that format \((p < 0.05)\).

As for the relative importance of the five entry decisions, we present the long-run elasticities computed at the mean in Table 4.\(^{21}\) In case of greenfield expansion, the order of entry is of crucial importance, with an elasticity (−0.70) two to three times as large as for scale (0.35), format adaptation (−0.23) and format familiarity (0.29). Faced with a trade-off between an early entry on a more limited scale or later entry on a large scale, companies following this international expansion route are better off to already pre-empt the locations they can. Indeed, one is forced to look for sub-optimal locations for new outlets once the best locations are taken. When dealing with acquisitions and joint ventures, the relative importance of being early is reduced (from −0.70 to −0.39). This result is not surprising, as with this mode of entry one maintains the option to acquire or join forces with earlier entrants that occupy the better locations, irrespective of one’s own entry timing. Hence, the competitive disadvantages of being late are much harder to reverse in the case of greenfield expansion than in the case of acquisitions and joint ventures. In terms of the relative importance of supply (familiarity to the parent company) and demand (newness to the target market) considerations...

\(^{21}\) These elasticities explicitly account for the impact of the interaction terms, as described in Hanssens et al. (1990, p. 42).
ence across the two regions, however, and the question remains whether some of the strategic entry decisions have a more pronounced, or rather an attenuated, effect when entering the transition economies of Eastern Europe. Arnold and Quelch (1998), for example, point out that firms entering emerging markets may need to rethink their marketing frameworks instead of using their “developed world” timing, partnering and format strategies as a default option, suggesting that also the slope parameters of the entry strategy decisions may differ. Tests on the equality of the individual main-effect parameters across both samples revealed no statistically significant differences in the impact of the order, scale and format adaptation decision. Differences were found, however, for the mode of entry and the format familiarity decisions. **23** First, the benefits of greenfield operations are even more pronounced in transition economies (\( p < 0.01 \)). This result may be attributed to the fact that incumbent retail operations tend to be worse in Eastern than in Western Europe. This may in turn result in higher transaction costs in emerging economies, suggesting a need for more hierarchical governance structures (Hoskisson et al., 2000). A second difference was found in the fact that using the same format as the parent firm had less of a positive impact in emerging markets (\( p < 0.01 \)). This may be due to the fact that there are less commonalities to be exploited when going from a Western to an Eastern European country than when staying within other Western European countries. These exploratory differences could provide interesting avenues for product development and reverse learning (Arnold and Quelch, 1998), but clearly need more detailed investigation than possible within the scope of the current study.

5.2. Efficiency

In terms of the impact of the strategic entry decisions on the subsidiaries’ asymptotic efficiency level, we again find all five strategic entry decisions to have a significant main effect. As was the case for sales performance, a positive impact is found for entry scale (\( p < 0.05 \)) and mode of entry (\( p < 0.01 \)), while a negative relationship is observed for order of entry (\( p < 0.01 \)). **24** Store formats that are new to the host market (\( p < 0.05 \)), and familiar to the parent firm (\( p < 0.05 \)), also improve long-term efficiency. In contrast to long-term sales performance, evidence of only one interaction effect was found, i.e. the positive effect of entering with a store format familiar to the parent firm is again reinforced when that format is more established in the target market (\( p < 0.1 \)). Hence, of the nine strategy-related effects, similar results are found for both performance measures in seven instances. In terms of the relative influence of the different strategy decisions, we again find that order of entry is the most critical dimension (with an elasticity of \(-0.24\)), even though the observed impact differences are no longer mode dependent.

Several interesting findings emerge for the control variables. As before, the retail-mix variables play a significant role in shaping long-term efficiency performance, but this time assortment size (\( p < 0.01 \)) and loyalty cards (\( p < 0.01 \)) turn out to have a positive and significant impact, while the choice of pricing strategy no longer has an impact. The latter finding corroborates previous findings that EDLP has volume advantages, but that these are insufficient to translate in higher profits (in our case, efficiency) when accounting for the efforts needed to achieve these volume gains (Kahn and McAlister, 1997).

With respect to the parent firm, we again find that their skills and resources contribute to the long-run performance level of their foreign operations. As with the sales variable, backing by a larger parent firm (\( p < 0.01 \)) and international experience (\( p < 0.01 \)) have a positive impact on long-run performance. Moreover, a large private-label share in the home market has, as expected, a negative influence on the efficiency of the foreign subsidiary (\( p < 0.05 \)). More inputs are therefore needed to achieve a given sales level for these firms, as they have to overcome a number of hurdles not faced by retailers relying more on national brands (see Section 2.2.2).

---

23 To avoid an over-parameterization, we tested this equality one strategy at a time on a model without interaction terms (i.e. no three-way interactions as \( \log(\text{order}) \times \log(\text{scale}) \times \text{trans} \) were considered).

24 Again, no evidence of higher-order effects was observed, neither in the main effect nor in any of the interaction effects.
insignificance of the private-label parameter in the sales equations suggests they are both willing and able to make this extra effort.

In terms of the host-market characteristics, we again find a significant positive impact of population size \( p < 0.01 \). Furthermore, we see that the competitive structure in the host market, while insignificant in terms of long-term sales performance, impacts the efficiency of the foreign subsidiaries: a significant negative impact is found for the number of retail competitors \( p < 0.01 \) and a positive significant effect for the variation in market shares \( p < 0.05 \). Finally, a significant negative impact for the transitional nature of the economy \( p < 0.05 \) is observed.

As for the sales-performance measure, we subsequently explored whether also the impact parameters of the entry-strategy decisions differed across both regions. Tests on the equality of the individual main-effect parameters revealed the same conclusions as before: the benefits of greenfield operations are more pronounced in transition economies \( p < 0.01 \), while the use of the same format as the parent firm had a smaller positive impact in those markets \( p < 0.01 \). No significant differences were, once more, observed for scale, order and format adaptation.

In sum, considerable consistency exists across both performance measures, both in terms of the impact of the strategic entry decisions and in terms of the moderating impact of the different control variables.

### 6. Validation

To assess the robustness of our findings, we conducted several validation checks. Specifically, we assessed the sensitivity of our results to (1) the length of the observation window, (2) the model

| Table 5 Validation results: sensitivity to the length of the observation window and the Gompertz specification |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Sales           | Efficiency       |                 |                 |                 |                 |                 |
| Strategic entry decisions | \( \beta_{\text{base}} \) | \( \beta_{\text{limited window}} \) | \( \beta_{\text{without recent}} \) | \( \beta_{\text{logistic}} \) | \( \beta_{\text{base}} \) | \( \beta_{\text{limited window}} \) | \( \beta_{\text{without recent}} \) | \( \beta_{\text{logistic}} \) |
| Scale of entry (\( \beta_1 \)) | 0.35\( ^a \) | 0.37\( ^a \) | 0.34\( ^a \) | 0.44\( ^a \) | 0.08\( ^b \) | 0.08\( ^b \) | 0.07\( ^b \) | 0.06\( ^b \) |
| Mode of entry (\( \beta_2 \)) | 0.96\( ^a \) | 1.08\( ^a \) | 1.00\( ^a \) | 0.64\( ^a \) | 0.41\( ^b \) | 0.48\( ^b \) | 0.47\( ^b \) | 0.23\( ^b \) |
| Order of entry (\( \beta_3 \)) | -0.39\( ^a \) | -0.37\( ^a \) | -0.42\( ^a \) | -0.21\( ^a \) | -0.24\( ^a \) | -0.21\( ^a \) | -0.26\( ^a \) | -0.13\( ^a \) |
| (Order of entry) \( \times \) | -0.04 | -0.06 | -0.04 | -0.06 | -0.04 | -0.03 | -0.03 | -0.07\( ^b \) |
| (Scale of entry) (\( \beta_4 \)) | -0.31\( ^d \) | -0.38\( ^e \) | -0.31\( ^d \) | -0.38\( ^d \) | -0.10 | -0.06 | -0.10 | -0.14\( ^d \) |
| (Mode of entry) (\( \beta_5 \)) | -0.26\( ^d \) | -0.34\( ^e \) | -0.24\( ^d \) | -0.10\( ^d \) | -0.10\( ^e \) | -0.08\( ^e \) | -0.06\( ^e \) | -0.05\( ^e \) |
| Format familiarity (\( \beta_6 \)) | 0.10\( ^b \) | 0.05\( ^b \) | 0.03\( ^b \) | 0.04\( ^b \) | 0.05\( ^b \) | 0.03\( ^b \) | 0.02\( ^b \) | 0.08\( ^b \) |
| (Format familiarity) \( \times \) | 0.17\( ^b \) | 0.08\( ^c \) | 0.07\( ^b \) | 0.05\( ^c \) | -0.01 | -0.07 | -0.01 | -0.05 |
| (Mode of entry) (\( \beta_7 \)) | 0.04\( ^b \) | 0.04\( ^c \) | 0.02\( ^b \) | 0.02\( ^b \) | 0.02\( ^c \) | 0.02\( ^c \) | 0.02\( ^c \) | 0.01\( ^c \) |

\( \beta_{\text{base}} \): refers to the parameters obtained in the original Gompertz model.
\( \beta_{\text{limited window}} \): refers to the parameters obtained when the observation window is limited.
\( \beta_{\text{without recent}} \): refers to the parameters obtained when the most recent entries are excluded.
\( \beta_{\text{logistic}} \): refers to the parameters obtained for the logistic model.

\( ^a \) \( p < 0.01 \) (one-sided).
\( ^b \) \( p < 0.05 \) (one-sided).
\( ^c \) \( p < 0.10 \) (one-sided).
\( ^d \) \( p < 0.01 \) (two-sided).
\( ^e \) \( p < 0.05 \) (two-sided).
\( ^f \) \( p < 0.10 \) (two-sided).
6.1. Sensitivity to the length of the observation window

Recent empirical and simulation evidence from Van den Bulte and Lilien (1997) shows that the estimated asymptotic performance may be sensitive to the last available observation. This was formally demonstrated in the context of the well-known Bass model, but similar patterns were expected in other non-linear specifications, such as the Gompertz model. To assess the robustness of our substantive findings to the length of the observation period, we re-estimated our model for a different end point, i.e. we used one observation less for every entry.

As shown in Table 5, our substantive conclusions on the long-run impact of the strategic-entry decisions are robust in terms of sign, magnitude and significance. For both sales and efficiency, we again find support for the long-run impact of all five entry decisions. Moreover, also with respect to the interaction effects, similar results are found.

A related issue deals with the sensitivity of the pooled estimation results to the more recent entries. On those entries, only a few data points are available, and the inflection point in their individual S-shaped growth patterns may not yet have been reached, making it more difficult to make reliable long-run inferences from these observations. This raises the question to what extent these entries influence the long-run performance estimates, and hence, our substantive results. In a next validation check, we dropped all information on the 21 entries which took place in the last two years. Our substantive findings on the long-run impact of the strategic entry variables were once more unaffected, both in terms of the main and the interaction effects.

6.2. Sensitivity to the choice of the Gompertz model

To ensure that our substantive findings are not idiosyncratic to the adopted (Gompertz) model specification, we compared our parameter estimates to those of an alternative growth-model specification. Specifically, we considered the logistic model (Hansens et al., 1990, p. 41). Making the market potential parameter time-varying and a multiplicative function of the entry decisions and control variables described before, a special case of the smooth transition regression model described in Granger and Terasvirta (1993, Section 4.2) is obtained.

For the sales equation, the same substantive conclusions as before are obtained, in that all five main effects are again significant (with the same sign as in the Gompertz model), while also the same three interaction effects are found to be significant. For the efficiency model, similar findings are obtained for seven of the nine considered effects. None of the differences involves a sign switch, however, but arise because of a significance switch for two of the interaction effects, in that they become significant in the logistic specification (see Table 5).

6.3. Sensitivity to the exogeneity assumption

In our model specification, we assumed the entry decisions to be exogenous. Lieberman and Montgomery (1988), however, have argued that these decisions reflect an intrinsic choice on the part of the firm that may be driven by its performance expectations (see Moore et al., 1991 for a detailed discus-

---

25 Because of space limitations, we focus our validation discussion on the impact of the strategic entry decisions (main and interaction effects). In all models, however, the control variables and dummy variables were included in the estimation.

26 While both the Gompertz and the logistic model allow for an S-shaped evolution towards a saturation level, there are some important differences between both specifications. In the logistic model, a maximum growth rate is obtained when current performance is at 50% of the asymptotic performance level, as opposed to 37% for the Gompertz curve. As the true maximum need not coincide with either one, it is appropriate to test the sensitivity of our findings to this model property. In addition, the rate of growth is symmetric around the inflection point for the logistic curve, but not for the Gompertz model (Chow, 1967). This asymmetry has been argued to make the Gompertz curve more appropriate for modeling new product or venture performance, as the period of rapidly increasing sales has been found to often be shorter than the period where sales converge to their saturation level (Franses, 1994a). The Gompertz curve is not only a priori more appealing than the logistic curve. We applied the empirical testing procedure described in Franses (1994b) to all individual entry series that had sufficient observations, and found the Gompertz curve to be preferred in a vast majority (> 75% for sales; > 85% for efficiency) of cases.
sion): based upon a comparison of resources and opportunities, the firm may form certain expectations about both post-entry performance and its dependence on certain entry decisions, and these expectations may guide the eventual selection of strategies. As such, certain unobservable determinants of the entrant’s performance (expectations) may no longer be independent of the entry decisions, causing a correlation between the latter and the error term of the estimation equation. In such a situation, the entry decision should be treated as endogenous rather than exogenous, and single-equation estimation techniques will lead to biased estimates.

Following Moore et al. (1991), we tested for the endogeneity of the entry decisions through the Hausman–Wu test. Specifically, in the test equation (i.e. Eq. (3)), we included both the potentially endogenous variables (i.e. the strategic decisions) and instruments for these variables, where the latter are derived as the forecasts from an auxiliary regression linking an individual entry decision to the control variables (retail-mix, firm and target-country characteristics) measured at the time of entry.²⁷ A $\chi^2$-test on the significance of these instruments then constitutes the exogeneity test. This test was implemented on a decision-by-decision basis (i.e. where the instruments were added both in the main and the interaction effects). None of the five tests revealed any violation of the assumed exogeneity of the decision variables (using a significance level of $p < 0.05$).

7. Conclusion

The grocery retailing industry seems to be infected by a highly contagious globalization fever which has spread rapidly all over Europe, and is now affecting major US players as well. Even though many retailers feel that going abroad is an appropriate response to the saturation and increasing competition in their home market, little consensus exists as to what strategy to adopt in the process. Even though entry decisions are likely to shape the competitive platform from which (hopefully enduring) competitive advantages may be gained, little is known on their impact on post-entry performance, especially in the longer run. Not only are retailers uncertain as to whether the hoped-for economies of scale through a large-scale presence will ever materialize, they also have little insight on when it is best to enter a given market, how much they should adapt their familiar store format to local tastes, or whether similar strategies will prove optimal in Eastern and Western European countries.

In this paper, we addressed these issues by investigating simultaneously the long-run performance consequences of five aspects of international expansion: the scale, mode and order of entry, the format adaptation to local market conditions, and the familiarity of the adopted format to the parent firm. In the process, we extended on previous work in several respects. First, by simultaneously considering an extensive set of choices made at entry, better insights into their relative impact are obtained. Second, by considering the asymptotic performance derived from $S$-shaped growth curves, we had an explicit focus on the decisions’ long-run performance consequences, and made our substantive conclusions time-subscript independent. Third, the scope of our data set, with entries in more than 20 Western- and Eastern-European markets, allowed us to extend the predominantly North-American knowledge base. Fourth, our study was not positioned in a consumer goods or industrial setting, but considered with the retailing industry one of today’s most important service industries.

Our main conclusion is that the strategic decisions made at entry continue to influence the foreign operations’ post-entry performance, both in terms of long-run sales and long-run efficiency. Especially the time (order) of entry appears to be a prime source of persistent competitive advantage, particularly when entering through greenfield expansion. Hence, firms still considering a potential entry towards “popular” destinations such as Poland or the Czech Republic, where over 20 foreign competitors have preceded them, should realize that they face a considerable barrier to long-run success because of this tardiness. It is fair to say that not just in those two countries, but actually in most European markets, the most attractive “pole positions” have been taken for some

²⁷ For the discrete mode-of-entry decision, a logit model was used as auxiliary regression; in all other instances, a multiplicative specification was adopted.
time. Later entrants will therefore have to take recourse to other (less effective) instruments to compensate for the ensuing disadvantage. Outside Europe, in contrast, the internationalization wave is just taking off. If our findings generalize to these cross-continental moves (cf. infra), a wait-and-see attitude is not recommendable. The main players seem to share this point of view, as they are recently rushing to supplement their European operations with activities in South America (e.g. Carrefour in Chile and Columbia or Ahold and Casino in Argentina) and Southeast Asia (e.g. Tesco in South Korea or Carrefour in Indonesia).

Second, companies often find themselves in a bind on which retail format to choose when entering a foreign market. This choice can be dictated by demand (adaptation to host-market conditions) or supply (expertise in the home market) considerations. Especially when entering through acquisitions or joint ventures, one should try to fully exploit the expertise of the local partners, and adapt the format to local market conditions, rather than trying to impose one’s own preferred format to achieve higher long-run sales.

Third, we also conducted exploratory analyses on whether one can transfer strategies that have been applied successfully when entering developed, Western-European, markets to entries into the transition economies of Eastern Europe. Even though all five entry strategies continue to have an impact on the long-run performance of the foreign operation, we find that the relative importance of some of them changes. For instance, the benefits of entering through greenfield expansion become more pronounced, while less benefits are incurred when entering with the parent’s most familiar format.

The current study has various limitations, which offer immediate avenues for future research. First, when testing the long-run impact of the strategic entry decisions, we controlled for a variety of retail-mix, firm, and host-market characteristics. These do not constitute an exhaustive set of post-entry performance drivers, however, and further research should identify other factors involved in the success of retailers going abroad. A second limitation deals with the operationalization of some of the variables included in our model. For example, the pricing strategy was operationalized through a single dummy variable (EDLP versus HiLo). Shankar and Bolton (1999) identified other relevant pricing dimensions (e.g., promotional intensity, price/promotion coordination, etc.), and observed substantial variability in pricing strategy across product lines. More intricate measures may therefore be called for. Similarly, we looked at non-monotonic effects for order of entry, but we could also envision non-monotonic effects for the other entry decisions such as scale of entry (Shankar, 1999). Also, order of entry was operationalized relative to the other foreign entries, thereby ignoring the number of major local incumbents one faces. While this is common practice in the internationalization literature (see, e.g., Mitchell et al., 1994) and the only feasible option in many transition economies (but less so in Western Europe), one may wonder to what extent this may have confounded our order-of-entry comparison across both trading zones.

As a final operationalization issue, it may be worthwhile to extend the set of performance indicators to profit and ROI measures. Third, apart from the demand-shift variable, all control variables were time-invariant, and measured either in the year preceding the entry (parent-firm and target-country characteristics), or observed in a given year following entry (retail mix). Obviously, these values may change over time, and affect the long-run performance potential. An extension of our approach would be to simultaneously control for their value at the time of entry (across-entry variation) and subsequent deviations from these initial values (within-entry variation), as described in Van den Bulte (2000).

Other areas for future research remain open as well. For instance, we focused attention on the long-run performance consequences of the strategic entry decisions. Apart from this “end point”, one may also want to study what drives the trajectory or evolution towards this long-run equilibrium, i.e. the impact of the entry decisions on both initial post-entry performance and subsequent growth. An extension of our modeling approach to jointly consider all three performance aspects could therefore be envisioned. Also, our data describe the second internationalization wave in the retailing industry, in which many firms no longer limit their international expansion to neighboring countries. Cross-continental moves, however, are a fairly recent, and not yet well documented, phenomenon. As such moves are expected to become
more prominent, it would pay to assess the generalizability of our empirical findings to this more global arena.

Acknowledgements

The authors are indebted to I. Geyskens, D.M. Hanssens, J.B. Steenkamp, P. Vanden Abeele, L. Warlop and three anonymous IJRM reviewers for useful comments.

References

The Economist, 1999. Shopping all over the world. The Economist


Perspectives on joint competitive advantages in buyer–supplier relationships

Sandy D. Jap *

Department of Marketing, Sloan School of Management, Massachusetts Institute of Technology, Room E56-317, 38 Memorial Drive, Cambridge, MA 02142-1307, USA

Abstract

A critical outcome of competitive strategy is the attainment of competitive advantages. Recently, there has been a growing recognition that such advantages may reside in the boundaries of a firm—via its relationships with outside organizations. However, there is little understanding regarding how such advantages are created, eroded, and preserved in such relationships. In this paper, I summarize the findings around competitive advantages from three studies, all of which involve longitudinal empirical tests of over 200 industrial buyers and their suppliers in a variety of industries. The collective results indicate that specialized investments facilitate the attainment of joint competitive advantages and these advantages are positively correlated with economic outcomes, organizational behavior, and expectations of continuity. Competitive advantages can also be eroded over time for buyers by suspicions of \textit{ex post} opportunism that arise within the course of the relationship. However, the detrimental effects of opportunism suspicions for both firms can be mitigated via the strategic use of various governance modes such as bilateral investments, goal congruence, and interpersonal trust. © 2001 Published by Elsevier Science B.V.

Keywords: Competitive advantages; Buyer–supplier relationships; Boundaries of the firm; Opportunism

1. Introduction

One of the key outcomes of competitive strategy is the attainment of competitive advantages—the resources or capabilities that enable a firm to compete more effectively in the marketplace. In the strategic management literature, the “Resource-Based View” (RBV) of the firm offers the explanation that these differences may be due to heterogeneity in differentiated or superior resources relative to competitors (Teece, 1980; Wernerfelt, 1984). More recently, there is a growing recognition that this principle is generalizable to the boundaries of the firm, in relationships with organizational buyers and suppliers. Particularly noteworthy is the work of Dyer (1996; Dyer and Singh, 1998) who argues the need for specialized supplier networks. These networks interrelate the use of idiosyncratic investments, knowledge-sharing processes, complementary capabilities and effective governance to create competitive advantages. However, one shortcoming of the research in this literature is that few studies specify how these factors interrelate to develop, maintain, and erode competitive advantages over time.
This raises several questions in the area of buyer–supplier relationships: How are competitive advantages jointly created? How are they eroded? And how are they preserved in relationships over time? These are the motivating questions in this research. In this paper, I identify a subset of critical factors highlighted in the marketing, RBV, and transaction cost economics (TCE) literatures that illuminate how buyers and suppliers jointly create, manage, and erode competitive advantages. This is accomplished by reviewing a series of three longitudinal studies that examine the risks and returns associated with specialized, collaborative relationships between firms (Jap, 1999; Jap and Anderson, 1999, 2000). All of these papers are generated from the same dataset, involving 200+ buyers and their respective suppliers. The longitudinal nature of the data enables the examination of causality issues, which are more often assumed than tested in past research.

These papers explain general performance; they do not focus on the attainment of joint competitive advantages, although this variable is embedded within the conceptual models of each paper. The purpose of the present paper is to highlight the specific results around this variable and consider positive spill-over effects on the relationship over time. Thus, this paper makes several contributions to our understanding of the dynamics of competitive advantages at the boundaries of the firm. First, it illuminates specific empirical results from various studies, consolidating these results within one paper to provide an overall perspective of how joint competitive advantages are developed, maintained, and eroded over time. Second, it supplements the insights from these papers with additional empirical analysis that suggests that the attainment of these advantages has a strong, positive relationship with changes in the dyad’s economic performance, collective functioning, and relationship stability over time. And finally, it provides a theoretical perspective on the conditions that facilitate such advantages at the boundaries of the firm.

The topic of competitive advantages at the boundaries of the firm is particularly timely, given recent growing interest in the “virtual firm” and the “extended enterprise”. These discussions view organizations as decentralized networks of financially independent organizations that are coordinated in such a way as to appear and behave as one unified organization. Understanding how competitive advantages are developed between each linkage has important ramifications for the overall functioning of the organizational network.

The structure of the paper is as follows. After a brief review of the conditions that facilitate the attainment of joint competitive advantages in buyer—supplier relationships, learnings from the three empirical papers are reviewed, as relevant to the development, erosion, and preservation of competitive advantages. This is followed by a discussion of the spillover effects of competitive advantages on buyer–seller relationships and a set of conclusions regarding key learnings. Data characteristics and details regarding empirical analyses are available in various appendices.

2. Conceptual framework

2.1. Literature review on competitive advantages

In an industrial supply context, competitive advantages are defined as \textit{strategic benefits gained over competing dyads that enable the dyad to compete more effectively in the marketplace} (Sethuraman et al., 1988). In the RBV framework, there are four theoretical conditions that underlie the achievement of competitive advantages: (i) resource heterogeneity, (ii) \textit{ex ante} limits to competition, (iii) \textit{ex post} limits to competition, or causal ambiguity, and (iv) imperfect mobility. These characteristics also create the backdrop for the attainment of competitive advantages in interorganizational relationships. In this section, these conditions are briefly reviewed and its relevance is discussed in an industrial supply context.

\textit{Resource heterogeneity} refers to the resource bundles and capabilities that underlie production in a firm (Barney, 1991). These resources have varying levels of productivity efficiency that enable firms to produce more economically or better satisfy customer demands than their competitors. When these factors are inelastic in supply and insufficient to satisfy demand, then the low-cost firm will earn supernormal profits in the form of rents to their scarce resources. Other high-cost firms will break even. This is known as the Ricardian rents argument.
(Ricardo, 1817; Rumelt, 1987). A key aspect of the argument is that the superior resources remain limited in supply. This allows efficient firms to sustain their competitive advantage as long as the resources cannot be expanded or freely imitated by competition. Prahalad and Hamel (1990) note that core competencies that are enhanced as they are applied (i.e., those which involve collective learning and are knowledge based) contain natural learning trajectories that also serve as a basis for competitive advantage.

In industrial supply relationships, buyers and suppliers bring together unique competencies in differing functional areas. These competencies may involve learning curves and differing levels of efficiency. When they are combined, the dyad gains access to critical resources that enable the creation of superior value in the marketplace. The more unique the combination of capabilities and the more inelastic the supply of the joint capability, the greater the potential for generating supernormal returns relative to competing dyads. For example, suppose a computer chip manufacturer and a car manufacturer consider the possibility of developing a chip that optimizes the performance of a new car. The chip manufacturer possesses chip design capabilities while the car manufacturer has car production capabilities. The heterogeneity of capabilities along with the inelasticity of supply of these capabilities allows for the creation of a unique, joint competency (i.e., car-chip optimized capabilities). This competency serves as the basis for superior product offerings to downstream customers.

Ex ante limitations to competition means that there must be limited competition for a particular resource position prior to any firm’s establishing the position. Barney (1986) contends that economic performance depends not only on the returns from various strategies, but also on the cost of implementing the strategies. Imperfections in strategic resource markets, where the necessary resources for implementation are acquired, enable the creation of supernormal returns. Without these imperfections in the markets, firms can only hope for normal returns. Rumelt (1987) argues that unless there is a difference in the ex post value of a venture and the ex ante cost of acquiring the necessary resources, the entrepreneurial rents are zero.

The chip-car collaboration between the two manufacturers may be an area in which there is limited competition. In the car market, there are many ways in which car manufacturers can create new customer value and earn supernormal rents. Some might offer better designs, lower costs, more efficient distribution, etc. Each car manufacturer may choose one or more of these bases by which to differentiate its competitive offering. Since each manufacturer is heterogeneous in its ability to exploit or access these resources, the potential for competitive advantages is created. In the previous example, the design of a unique microprocessor that optimizes the performance of a new car is just one of many possible ways in which a car manufacturer may compete in the industry; hence, competition for this position may be limited or even unrecognized as a possible basis for competitive strategy. In order for competitors to duplicate this advantage, they would have to collaborate with the chip manufacturer or some other chip manufacturer and have similar capabilities for exploiting the joint competency.

Ex post limitations to competition is the notion that once a firm is able to gain a superior position, there must be barriers to competition for the associated rents. The work in this area has focused on two forms of ex post competition: imperfect substitutability and imperfect imitability (Barney, 1991; Reed and DeFillipi, 1990). Substitutes erode rents by making demand more elastic. Much greater attention has been paid to the concept of imperfect imitability. This is the notion that competitors have difficulty imitating the resource stream and eroding the firm’s rents. Lippman and Rumelt (1982) call this ‘causal ambiguity’, which prevents would-be imitators from knowing exactly what to imitate or how to go about it. Other similar mechanisms include producer learning, switching costs, reputation, search costs, or other barriers to entry.

Relationships between organizations are particularly amenable to ex post limitations to competition because it is very difficult for competitors to observe and duplicate the efforts and activities of the dyad. Returning to the car-chip manufacturer example, the development of a unique microprocessor is an elaborate process involving considerable specific, tacit, and complex information. In order for competitors to erode the rents that result from these efforts, the
competition must be able to observe and easily duplicate the dyad’s interactions and collective functioning. They must understand the ‘causal’ structure of activities, communication, and work processes occurring within the dyad that creates the basis for competitive advantage. This is a very difficult thing to accomplish.

Imperfect mobility refers to resources that are not easily traded; they are more valuable within the firm than in other firms. This includes resources that are ill defined or non-fungible. Ill-defined resources are those for which there are no well-defined property rights. Examples of this may include customer loyalty or supplier trust. Non-fungible resources are those whose value is derived within a specific context; its value is not transferable to alternative relationships or firms. These resources may be tangible (e.g., capital equipment, manufacturing facilities), or intangible (e.g., human resource capabilities, specific technologies and know-how). Teece (1986) describes co-specialized resources as another case in point. These are resources that have higher value when employed together than when employed separately. When co-specialized or non-fungible resources have few other equivalent uses or value outside the firm, then they are imperfectly mobile. These resources have less value outside the firm, hence, they are not readily bid away, remaining bound to the firm and available for use over the long run.1 This provides a basis for competitive advantage.

In the chip-car manufacturer example, the unique microprocessor that is developed between them is a non-fungible, co-specialized resource that enables superior value over competitive offerings. In order for competitors to generate the same offering, they must work with the chip manufacturer or a competing chip manufacturer to develop another unique microprocessor that is optimized for their cars. This can take considerable time, energy, and effort to accomplish and may ultimately be imperfect. In the meantime, the rents of the original chip-car dyad and unique microprocessor are preserved.

In sum, industrial supply relationships are particularly amenable to the generation of competitive advantages, because (i) the buyers and suppliers are heterogeneous in their resources and capabilities, (ii) they can identify joint positions for which there may be limited competition, (iii) their activities are difficult for competitors to observe and duplicate, and (iv) the creation of specialized, idiosyncratic investments between them enables a distinct advantage over competitors that can be realized into the long term. By moving away from arms-length exchanges and specializing their relationships through idiosyncratic investments, knowledge exchange, complementary competencies, and more effective governance mechanisms, buyers and suppliers can create the potential for earning competitive advantages (Dyer, 1996; Dyer and Singh, 1998).

There are many examples of this in the marketplace. Baxter Healthcare works closely with hospitals on storing and distributing supplies to and within a hospital. By focusing on the creation of optimal inventory levels and non-price benefits, both parties receive several valuable outcomes—the hospital gains reduced inventory investment and operating costs, while Baxter receives increased revenue, market share and customer loyalty. Similarly, in 1986, Xerox worked closely with their suppliers to develop customized processes and components that reduced their copier manufacturing costs 30–40%. In turn, the suppliers received sales and volume guarantees, an enhanced understanding of their customer’s needs, and a strong position with Xerox for future sales.

The preceding discussion identifies the conditions under which buyers and suppliers might jointly create competitive advantages. However, it does not identify the processes by which organizations produce these advantages together, or the means by which these advantages are either eroded or preserved at the boundaries of the firm. In the following sections, we review three papers that test possible explanations for how advantages are developed and managed over time. The creation of competitive advantages is formulated using an RBV perspective, while the erosion and preservation of these advantages is based on a TCE explanation. Clearly, these explanations are not exhaustive of the many ways in which competitive advantages evolve over time, but they represent an initial step in understanding this

---

1 The disadvantage to such resources is that they may result in a reduction in flexibility in the face of environmental or technological changes.
complex phenomenon. The focus of this discussion is on the substantive results and implications of each paper, as analysis details and theoretical frameworks can be found in the specific papers. Collectively, these papers inform our understanding of a key outcome of competitive strategy, the central topic of this special issue. Details of the empirical methodology are contained in Appendix A.

2.2. Creating competitive advantages

Despite the widespread interest in the use of close, collaborative relationships, there is little understanding of how strategic outcomes are achieved at the boundaries of the firm. Dyer and his colleagues suggest that buyers and suppliers utilize complementary competencies, specialized investments, knowledge exchange processes, and various governance modes to create advantages. However, they do not specify how these factors interrelate to build these advantages. Recent work in marketing tests a possible set of interrelationships between Dyer’s four factors (Jap, 1999). Essentially, this research suggests that complementary competencies, environmental conditions, and human resources provide incentives for the dyad to create knowledge exchange processes, or coordination efforts, and specialized, investments. These two factors—coordination efforts and bilateral idiosyncratic investments—are critical for generating competitive advantages and economic performance. An overview of this conceptual model and estimation effects is depicted in Fig. 1.

When buyers and suppliers strive to attain competitive advantages, they must creatively consider how to leverage their unique environment, respective competencies, and human resources available to them. They recognize and understand that each firm’s success depends in part on the other firm. As such, their coordination efforts—the regular pattern of similar or complementary actions and activities (Anderson and Narus, 1990)—enable them to share information, opportunities, and processes in such a way as to facilitate the achievement of competitive advantages. These efforts may be manifested in the formation of joint projects tailored to the dyad’s

![Fig. 1. Creating competitive advantages. The italics indicate that the data were collected one year later. Intercorrelations, factor loadings, and measurement and latent errors are not included for simplicity of depiction. The measurement model for buyers has a $\chi^2$ of 1028.9 (404 df, $p < 0.01$), a comparative fit index (CFI) and incremental fit index (IFI) of 0.87 and a Tucker–Lewis Index (TLI) of 0.85. The root mean square error of approximation (RMSEA) is 0.075. The $\chi^2$ for the suppliers’ measurement model is 704.6 (404 df, $p < 0.01$), with CFI = 0.91, IFI = 0.91, and TLI = 0.89. The RMSEA = 0.058. The $\chi^2$ of the two-group structural model is 1689.31 ($p < 0.01$) with 784 df, CFI = 0.88, IFI = 0.88, TLI = 0.87, and RMSEA = 0.048.](image-url)
needs and an ongoing effort to exploit existing synergies and unique opportunities between the firms. Although the empirical results suggest that these efforts have a marginally positive effect ($\beta = 0.11$, $p < 0.10$) on the achievement of competitive advantages one year later, the directional consistency is informative.

Idiosyncratic investments are non-fungible investments that uniquely support the buyer–supplier relationship (Williamson, 1985). The joint creation of these investments promotes the achievement of strategic outcomes because they promise efficiencies in coordination and promote interfirm cooperation and enhanced performance (Dyer, 1996). The empirical results verify this, indicating that such investments do play a significant role ($\beta = 0.35$, $p < 0.01$) in the buyer and supplier’s ability to realize joint competitive advantages up to one year later. This finding is consonant with recent studies that indicate that productivity gains are attainable when the parties to an exchange are willing to make idiosyncratic investments and combine their resources in unique ways (Asanuma, 1989; Dyer, 1996). Amit and Schoemaker (1993) argue that specialization of investments is “a necessary condition for rent”, and that “strategic assets by their very nature are specialized”. Hence, by definition, buyers and suppliers must do something specialized or unique to develop a competitive advantage.

Collectively, these results suggest that the use of bilateral idiosyncratic investments may enable the achievement of competitive advantages via lower total value chain costs, enhanced product differentiation, fewer defects, and faster product development cycles. Differentiation, however it is achieved, is a necessary condition for competitive advantages. The interaction patterns and coordination efforts transpiring across the dyad, along with dedicated investments, create the differential advantages that the firms strive for together. This general result echoes Harrigan’s (1986) study of joint ventures. She found that interorganizational strategies for competitive or strategic purposes typically required substantial re-

2 Competitive purposes may include preemption of competitors and influence of industry structure, while strategic purposes include activities such as new market entry, technology transfer, production synergies, etc.

sources and coordination effort. It is the combination of the two—effort and investments—that provides the foundation for the inimitable aspects of the collaboration process as a source of strategic outcomes.

2.3. Eroding competitive advantages

Increasing specialization in interfirm relationships is not without cost. By making specific commitments in effort and investments to a particular strategy, the dyad will forego alternative opportunities. They might also be less flexible to technological changes in their environment. The non-fungible nature of idiosyncratic investments also creates the potential for “hold-up” problems and additional forms of opportunism.

Opportunism is self-interest seeking with guile, and includes overt behaviors such as lying, cheating and stealing, as well as subtle behaviors such as dishonoring an implicit contract, shirking, failing to fulfill promises, and obligations. It is the equivalent of bad faith, the implication being that the party who is opportunistic is not trustworthy. In an industrial supply setting, opportunism may involve misrepresentations (e.g., making hollow promises or “window-dressing” one’s efforts), unresponsiveness (e.g., aloofness), unreasonable demands (e.g., asking the other party to pay more than their fair share of a problem), and lying.

The notion of opportunism is what differentiates TCE from alternative conceptualizations of the firm, such as agency theory, relational exchange theory, or RBV. The TCE presumption is that economic actors attempt to forecast the potential for opportunism as a function of unfolding circumstances, then take preventive actions in transactions where opportunism is likely to be high. Opportunism is an explanatory mechanism, not readily observable, and typically empirically untested. However, it is important be-

---

3 This aspect of TCE has been criticized, most notably by Ghoshal and Moran (1996), who argue that TCE fails to consider that agents can behave in both opportunistic and trustworthy ways. Hence, the TCE premise renders a distorted view of reality. Commercial life is comprised of both opportunistic and trustworthy behaviors. In fact, trustworthy behavior is both pervasive and of paramount importance for the functioning of the firm in an economic system.
cause it has potential for enormous impact on economic performance. The work to date has, instead, focused on the antecedents of opportunism (Anderson, 1988; John, 1984; Smith and Barclay, 1997) and strategies for minimizing opportunism (Brown et al., 2000; Wathne and Heide, 2000), but not its impact on the performance and collective functioning of exchange relationships.

Recent research considers the impact of opportunism on a number of outcomes of buyer-supplier exchanges, ranging from the collaborative tendencies and problem-solving orientation of the dyad to economic performance, long-term expectations, and the realization of joint competitive advantages (Jap and Anderson, 1999). The conceptual model and estimation results are illustrated in Fig. 2. In this research, we find that the mere suspicion of opportunism is sufficient to damage, even destroy, a relationship, regardless of whether the suspicion is factually justified (cf., Ross et al., 1997).

As a buyer or supplier becomes increasingly suspicious of its counterpart, there is a decreasing motivation to make the exchange successful through coordination efforts and the creation of bilateral idiosyncratic investments. These activities typically require the sharing of sensitive information in order to identify opportunities for joint success. If one of the parties in the exchange suspects that the counterpart is acting opportunistically, the firm will be reluctant to share such information, thus adversely affecting the exchange activities and investments necessary for achieving competitive advantages, and exchange activities. Organizational theorists have observed that when exchange deteriorates, behaviors such as tolerance of incompetence, replacement of substance with form, scarcity of clear goals and decision benchmarks, and loss of effective communication, are evidence of some of the first observable signals of decline (Lorange and Nelson, 1987).

As interfirm coordination processes disintegrate, there is diminishing incentive for the firms to continue to specialize their relationship through idiosyncratic investments because the ability to exploit these investments has deteriorated. In this manner, opportunism suspicions across the dyad can cause the exchange relationship to take a dysfunctional turn. Hence, over time, one would expect that opportunism suspicions would facilitate the erosion of competitive advantages. This is empirically verified for buyers ($\beta = -0.33, p < 0.05$), but not suppliers ($\beta = -0.11, \text{ns}$).

Why do opportunism suspicions erode the buyer’s perspective on joint competitive advantages but not the supplier’s perspective? This may be due to issues of distributive justice that are not captured in the model. For example, it may be that the competitive advantages, although jointly attained and created, may not be equally shared or valued between the firms. If the buyer receives 60% of the benefits from creating competitive advantages and the supplier receives 40%, the buyer may be more likely to focus on the erosion of competitive advantages, whereas the supplier may not be as concerned because it still receives a substantial portion of the benefits.

![Fig. 2. Eroding competitive advantages. The italics indicate that the data were collected from suppliers. Measurement model for buyers and suppliers pooled: $\chi^2$ of 1671.2 (695 df, $p < 0$), CFI = 0.92, IFI = 0.92, TLI = 0.90, RMSEA = 0.053.](image-url)
ceives 40%, buyers may feel more threatened by opportunism suspicions than suppliers. This is not uncommon in industrial procurement exchanges. Suppliers generally serve their customers, even if the exchange contains bilateral dependence and relatively close ties. While the overall results of this research indicate that opportunism suspicions generally have a negative effect on outcomes for both buyers and suppliers, the impact on joint competitive advantages is differential across the dyad. Additional research is needed to better inform our understanding of how and why this occurs.

2.4. Preserving competitive advantages

Because of the many circumstances within buyer–supplier relationships that permit opportunism, TCE recommends that a governance structure be put in place as a means to reduce it, since total elimination is typically infeasible or prohibitively expensive. Toward this end, considerable research in various disciplines has been directed toward the use of governance mechanisms as a means of reducing the incentive for opportunistic behavior. Within marketing, numerous studies have considered the antecedents of various mechanisms, including pledges in the form of bilateral investments, explicit and normative contracts, relational norms, qualification procedures, and monitoring. Given this, it is surprising that there are virtually no studies that consider whether these mechanisms are actually capable of safeguarding the firm’s critical outcomes. All of these mechanisms impose varying economic or social costs of opportunistic behavior and are typically put in place in the relationship as a preventative measure. However, over time, circumstances do change and ex post opportunism suspicions may arise as a result of the change in circumstances. When this happens, it is not clear whether the governance mechanisms that were put in place earlier in the relationship are able to safeguard the outcomes of the relationship against the poisonous effects of opportunism ex post.

A recent paper considers the safeguarding capability of three different governance mechanisms in the presence of ex post opportunism (Jap and Anderson, 2000): bilateral idiosyncratic investments, goal congruence, and interpersonal trust. These three mechanisms have distinct theoretical roots and characteristics; all have received considerable attention in the literature on interorganizational relations. Bilateral idiosyncratic investments figure prominently in TCE, while goal congruence is central to agency theory. Both are interorganizational properties. Interpersonal trust is a property of the boundary-spanning individuals; it is a construct often noted in the growing literature on relationship marketing. These three theoretical perspectives—TCE, agency theory, and relationship marketing—have had an enormous impact on research in the channel management literature.

The safeguarding capability of these three governance mechanisms at time 2 is considered at varying levels of ex post opportunism (high and low) at time 1. An overview of the conceptual model and estimated effects is provided in Fig. 3. I now consider each mechanism in turn.

Bilateral investments. Earlier, I noted that bilateral investments are often made for the creation of value and the achievement of competitive advan-

![Fig. 3. Preserving competitive advantages. H refers to high ex post opportunism, L refers to low ex post opportunism. The italics indicate that the data were collected one year later. Intercorrelations, factor loadings, and measurement and latent errors are not included for simplicity of depiction. Measurement model for buyers and suppliers pooled: $\chi^2 = 969.54$ (377 df, $p < 0.001$), CFI = 0.94, IFI = 0.94, TLI = 0.93, RMSEA = 0.056. Structural model: $\chi^2 = 709.3$ (383 df, $p < 0.001$), CFI = 0.92, IFI = 0.92, TLI = 0.90, RMSEA = 0.052.](image)
tages, as mentioned earlier. However, these investments also have an important relationship stabilizing quality in that they represent a credible sign of each party’s commitment to the relationship (Anderson and Weitz, 1992) and result in a situation of “mutual hostage taking” that prevents both parties from behaving opportunistically. By “tying each other’s hands”, the parties can paradoxically strengthen their relationship (Schelling, 1960). Hence, such investments facilitate expectations of continued exchange into the future.

The results suggest that such investments are capable of safeguarding the achievement of competitive advantages against varying levels of ex post opportunism ($\beta = 0.55$, $p < 0.05$). When a firm begins to suspect its counterpart of opportunistic behavior, bilateral investments can shift the focus to mutual cooperation and create an incentive to maintain and continue the relationship until the value of its investments is recouped. By imposing costs on premature termination of the relationship, these investments enable the preservation of valuable outcomes, such as competitive advantages, in long-term exchanges. Even if higher levels of opportunism suspicions exist within the dyad, these investments continue to have a powerful, positive effect on the attainment of competitive advantages. Over time, subsequent investments may be even more connected, making the outcomes of each party increasingly interdependent (and increasing the safeguarding capability of bilateral investments).

Goal congruence. The domain of agency theory is to determine the most efficient contract to govern the relationship, given that the parties are self-interested, bounded in rationality, risk-averse, and having goal conflict (see Bergen et al., 1992 and Eisenhardt, 1989 for a review). Agency theory notes that whenever cooperating parties have differing division of labor, goal conflicts can create incentives for opportunistic behavior (Jensen and Meckling, 1976). The goals of exchange for buyers and suppliers are typically at odds with each other. Buyers desire the most cost-effective purchase, while suppliers strive to achieve the sale with the highest profit margins or revenue potential.

By developing goal congruence—the extent to which firms perceive the possibility of common goal accomplishment (Eliashberg and Michie, 1984)—between the parties, the incentive for opportunism can be curbed. Anderson (1988) finds that goal congruence is related in a convex way to opportunism; the more salespeople perceive alignment between their goals and the company’s goals, the less opportunism they practice on the job—and at an increasing rate. By developing common goals between the buyer and supplier, a perception is created that what is beneficial for the counterpart will also be in the best interests of the firm. This creates a reduced incentive to act opportunistically and inhibits the development of suspicions within the dyad. It also incents the dyad to seek mutual gains and forego individual gains detrimental to their joint returns.

However, the empirical results are interesting in that they suggest that goal congruence has its greatest impact on the preservation of competitive advantages when the threat of ex post opportunism suspicions is high ($\beta = 0.34$, $p < 0.05$). When suspicions are low, goal congruence has no effect ($\beta = -0.21$, ns). Evidently, greater goal congruence becomes a more effective way to keep the relationship on track for achieving strategic outcomes and bolsters confidence in the long-term future of the arrangement. In contrast, when things are running smoothly (very little opportunism is suspected), higher levels of goal congruence have a lesser impact on the relationship’s outcome. This is an important result, as it highlights the differential value of goal congruence, depending on the nature of the relationship.

Why should goal congruence be a less valuable safeguard when things are going well? A likely explanation is that agreement on goals, on objectives, is taken for granted in ongoing relationships when opportunism is not suspected (Neilsen and Rao, 1987). This is because goal congruence is useful for aligning the firms’ incentives and charting the dyad’s course of activities in the early and buildup phases of a relationship. As the relationship reaches maturity, goal congruence may assume an unspoken or unarticulated quality, as the dyad fo-

---

4 Empirical results are supportive of the positive link between bilateral investments, commitment intentions, and joint activities. Gundlach et al. (1995) found a positive relationship between joint inputs and long-term commitment intentions. Zaheer and Venkatraman (1995) show a strong correlation between reciprocal investments and joint action in the area of new product launches.
cases on day-to-day activities and the ongoing operations of the exchange. However, as a firm comes to suspect that its counterpart is engaged in opportunistic behavior, goal congruence, where it exists, may be invoked as a means by which to evaluate and understand the counterpart’s deviation in behavior. Hence, goal congruence comes to the foreground, becoming salient. The firms may find that an appeal to common objectives (super ordinate goals) is a more effective way to resolve their differences and enhance the day-to-day functioning of ongoing exchanges between organizations, such that they continue to relate in a manner consistent with their shared goals. This improves the buyer and supplier’s ability to generate performance outcomes and extends its time horizon, thus preserving the competitive advantages jointly achieved.

**Interpersonal trust.** One of the most widely acknowledged social norms for governing and coordinating interorganizational exchange is trust (Morgan and Hunt, 1994). Trust is the ability to reliably predict the actions of the other party in the relationship and the belief that the other party will not act opportunistically if given the chance to do so (Andaleeb, 1992; Anderson and Narus, 1990). Increasingly, channel relationship researchers are acknowledging the impact of interpersonal factors such as trust, on interfirm outcomes. Larson (1992) found that personal relationships and reputations, coupled with knowledge of the firm’s skills and capabilities, shaped the context for new exchanges between firms by reducing risks and uncertainties about the motives and intentions of the other firm.

Several studies suggest that interpersonal trust operates in an independent, yet complementary manner to many organizational variables; that is, it facilitates relational processes, such as collaboration and relational norms, but has limited impact on performance (Smith and Barclay, 1997; Jap, 1999; Moorman et al., 1992). Yet the empirical results suggest that interpersonal trust is capable of safeguarding joint competitive advantages against varying levels of ex post opportunism ($\beta = 0.37, p < 0.05$). These results further suggest that the poisonous effect of opportunism suspicions may be limited to less tangible relational outcomes—such as expectations of continuity and evaluations of an exchange counterpart. Evidently, the tangible aspects of competitive advantages bolster them unsusceptible to changes in the interpersonal relationships between individual boundary spanners. It may be that such advantages have direct implications for economic performance or critical aspects like market strategy such that they are not quickly eroded by ex post opportunistic behavior. Hopefully, this finding will stimulate additional research to further understand the boundary conditions around this particular governance mode.

Collectively, the implication for managers regarding all three governance modes—bilateral investments, goal congruence, and trust—is that they do play important, but differing roles for safeguarding competitive advantages in the interorganizational relationship. Each is useful for safeguarding the achievement of competitive advantages, in the face of ex post opportunism, although goal congruence is less effective as ex post opportunism suspicions decrease. The impact of bilateral investments and interpersonal trust remain robust against varying levels of increased opportunism suspicions.

### 2.5. Spill-over effects of competitive advantages

The preceding discussion focuses on the dynamics of competitive advantage—how it is developed, eroded, and preserved over time. Another useful aspect to consider regards the potential implications of realizing competitive advantage on the relationship—the positive spill-over effects that occur with the attainment of competitive advantages over time. For example, it would be interesting to know whether the attainment of joint competitive advantages over time is accompanied by an improvement in economic outcomes or collective functioning as well.

To date, the literature indicates that long-term, collaborative relationships yield significant financial payoffs for both buyers and suppliers (Kalwani and Narayandas, 1995; Noordewier et al., 1990). Specifically, these studies show that these purchasing arrangements enable reduced possession and acquisition costs and enhanced inventory utilization, without sacrificing sales growth. This is accomplished via idiosyncratic investments, repeat sales and cross-selling opportunities, idea-sharing, switching barriers, and the use of supportive norms (e.g., flexibility, assistance, information exchange, etc.). However, these studies do not consider other strategic or
less tangible aspects. I address this gap by considering how improved financial performance over time is related to the achievement of competitive advantages and improvements in relationship functioning. I do not delve into the specific activities that create these outcomes, but instead examine the interrelationships between these outcomes using the same longitudinal data set of the previous three studies.

**Profit performance** is a perceptual measure of profits resulting from the dyad’s effort (the information expressed in currency was judged as too sensitive); it is a measure of the economic payoffs from collaborating together. The firm’s *expectations of relationship continuity* speak to the long-term viability of the relationship. *Constructive responses to problems* are a process measure indicating how it approaches and deals with problems that arise in the course of the relationship. Collectively, these variables tap a variety of critical aspects of buyer–supplier exchange: current financial performance, future expectations, and ongoing relationship functioning.

My expectation is that the realization of competitive advantages should be positively correlated with improvements in performance, continuity, and problem-handling behaviors because the achievement of competitive advantages may enable the firms to attain positions with higher rent-earning potential, and motivates them to continue working together into the future in a constructive manner so as to preserve and realize the potential rent earnings from this improved position. Hence, these factors reinforce each other over time; constructive communication and work processes enable the achievement of competitive advantages and financial performance, which bolsters their confidence in each other and provides new opportunities for the dyad to develop and enhance further. As the dyad grows in its understanding of each other, they are able to engage in more sophisticated planning activities and processes that may lead to greater positions of rent-earning potential. All of these uphold a variety of tangible and intangible performance outcomes.

I explore this possibility by correlating a measure of change in competitive advantages over a one year period, with changes in profit performance, expectations of continuity, and constructive responses to problems over time. The residual change score reflects the change from time 1 to time 2, while controlling for the impact of the initial starting point (time 1). In other words, these scores reflect the amount of change in a construct at time 2 per unit increase at time 1. Additional details are provided in Appendix B. The correlations of residual change scores for competitive advantages are: 0.53 with profit performance, 0.61 with expectations of continuity, and 0.49 with the focal firm’s response to problems. All of these are significant at \( \alpha = 0.001 \).

These results suggest that as joint competitive advantages are achieved over time, there is a corresponding improvement in profit performance, expectations regarding the future viability of the relationship, and interparty functioning. Similarly, if competitive advantages are eroded over time there should be a corresponding decrement in profit performance, future expectations, and constructive responses to problems. While this does not imply causality among the variables, it does suggest that when competitive advantages are developed in these relationships, there is a positive spillover effect on other critical aspects of the relationship: economic performance, problem resolution processes, and long-term expectations. The implication is that carefully managing the development of competitive advantages has important ramifications for the state of the overall relationship. There is also great value in monitoring the circumstances that lead to erosion of competitive advantages, as this will likely imply erosion in the overall state of the relationship.

### 3. Conclusions

This paper has reviewed the theoretical conditions that are foundational for the attainment of competitive advantages and has compiled specific results of various studies that illuminate how such advantages might be obtained, eroded, and preserved in buyer–supplier relationships. Clearly, the development of idiosyncratic investments plays a critical role, both in facilitating the achievement of joint advantages as well as preserving such advantages from *ex post* suspicions of opportunism. Interorganizational goal congruence and the trusting relationships of individuals also aid in the preservation and sustainability of these advantages against the poisonous effects of opportunism suspicions over time. This research also considers the spill-over effects of creating such ad-
vantages at the boundaries of the firm. These results indicate that the achievement of joint competitive advantages over time is positively associated with corresponding improvements in profitability, future expectations, and relationship functioning. Hence, the payoffs of achieving competitive advantages go well beyond the attainment of the advantages themselves, extending to other key aspects of buyer–supplier relationships.

Although this research is not exhaustive in nature, it does examine a subset of possible variables and represents an incremental step toward better understanding the complex phenomenon of how joint, competitive advantages are created in ongoing industrial supply relationships. The intent of this work is to stimulate additional research on this important aspect of competitive strategy. Future research ought to consider additional conditions for competitive advantage that are omitted here, such as environmental conditions and competitive actions. How buyer–supplier dyads respond to the competitive signals and actions of competing dyads is an avenue of inquiry never before considered. Additionally, the achievement of competitive advantages in electronic contexts is an important area of future inquiry. How does technology provide opportunities for bolstering or impeding competitive advantages? What can buyers and suppliers do together to leverage emerging technologies at the boundaries of the firm? The current market environment provides a ripe opportunity for considering such issues.

Acknowledgements

This research was funded by grants from the Alden G. Clayton Dissertation Competition and the Marketing Science Institute. Data support was provided by the Institute for the Study of Business Markets at the Pennsylvania State University. Many thanks to the co-editors, the anonymous IJRM reviewers and participants of the 1999 Conference on Competition and Marketing for helpful comments.

Appendix A. Methodology

All of the models of Figs. 1–3 are estimated on the same dataset, a longitudinal survey of over 200 buyer–supplier dyads in a variety of industries. In this section, I briefly describe the characteristics of the data and estimation procedures. Additional details can be found in each of the three papers (Jap, 1999; Jap and Anderson, 1999, 2000).

A.1. Data collection and sample

A.1.1. Procedure

The procurement divisions of four Fortune 50 manufacturing companies are represented in the dataset: a computer (PC) manufacturer, a photography equipment manufacturer, a chemical manufacturer, and a brewery. In order to maximize the sample size and minimize potential attrition effects at time 2, two hundred buyers from across the four firms were asked to report on two different supply relationships, in reference to a specific supplier and identify a knowledgeable, key informant at the supplier firm with whom it was currently working with and had experienced at least one year of frequent interaction. When the buyer surveys were returned, a parallel survey was sent to individuals in the supplier firms identifying the buyer firm and individual respondent and instructing the supplier to complete all items and questions in reference to the buyer firm. Thus, both the buyer and supplier used the other as a reference point for their respective surveys, such that the data collected reflect the same relationship as viewed by each side. One year later, the buyer and supplier were asked to complete similar surveys with respect to the same firm. The respondents completed identically worded, multiple-item, 7-point Likert scale measures with respect to the current state of the relationship. A listing of the scale items and reliabilities is presented in Appendix C. Construct means, standard deviations, and correlations are presented in Table A.1.

A.1.2. Sample characteristics

Two hundred seventy-five buyer surveys were returned at time 1 (a 69% response rate) and from these, 220 corresponding supplier surveys were completed (an 80% response rate), creating 220 matched pairs of dyads. At time 2, 167 buyer surveys and 154 supplier surveys were returned, representing a 61% and 70% response rate at time 2 among buyers and suppliers, respectively. The dyads had worked with
Table A.1
Means, standard deviations (SD), and correlations

<table>
<thead>
<tr>
<th>Construct</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Coordination efforts</td>
<td>5.1</td>
<td>1.1</td>
<td>1.0</td>
<td>7.0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Bilateral idiosyncratic investments</td>
<td>5.2</td>
<td>1.2</td>
<td>1.0</td>
<td>7.0</td>
<td>0.46</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Opportunism suspicions</td>
<td>2.2</td>
<td>1.0</td>
<td>1.0</td>
<td>6.6</td>
<td>–</td>
<td>0.43</td>
<td>0.09</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Goal congruence</td>
<td>5.1</td>
<td>1.1</td>
<td>1.3</td>
<td>7.0</td>
<td>0.63</td>
<td>0.31</td>
<td>0.56</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Interpersonal trust</td>
<td>5.8</td>
<td>1.0</td>
<td>1.8</td>
<td>7.0</td>
<td>0.41</td>
<td>0.27</td>
<td>0.54</td>
<td>0.49</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Realized competitive advantages</td>
<td>5.1</td>
<td>0.8</td>
<td>1.5</td>
<td>7.0</td>
<td>0.02</td>
<td>0.14</td>
<td>0.08</td>
<td>0.00</td>
<td>0.02</td>
<td>–</td>
</tr>
</tbody>
</table>

All correlations > 0.1 are significant at $\alpha = 0.01$.

Each other on an average of 3.7 years, indicating that they had a significant base of past history and transaction experience. Annual transactions involved over $63 million in materials and services, such as capital equipment, components, services, and maintenance, repair and operating supplies.

A.2. Analysis

A.2.1. Measurement

The unidimensionality and convergent and discriminant validity of the constructs in each figure was assessed via latent variable confirmatory factor analysis (CFA) techniques. Estimation is accomplished via full-information maximum-likelihood in LISREL 8.03 (Jöreskog and Sörbom, 1993). The specific fits of the models are displayed below each figure. Collectively, they demonstrate adequate fit, convergent validity and discriminant validity among each factor in the model.

A.2.2. Structural model estimation

The model of Fig. 1 is simultaneously estimated among buyers and suppliers using maximum-likelihood estimation of latent variables in LISREL 8.03; there were no significant differences between the $\beta$ parameter estimates across the two groups. The fit indices of this model are displayed below the figure.

The impact of opportunism suspicions on the realization of competitive advantages as shown in Fig. 2 is examined via an ordinary least-squares regression in which the impact of opportunism suspicions and control variables (i.e., the level of mutual dependence between the buyer and supplier, and purchase type) are estimated on competitive advantages one year later. The model specification for this regression is as follows: $Y = \beta_0 X_0 + \beta_1 X_1 + \epsilon$, where $Y$ represents the achievement of competitive advantages. $X_0$ is the intercept, and $X_1$ represents opportunism suspicions of the counterpart. $\beta_1$ is the regression coefficient, and $\epsilon$ is the random error term for the equation. The equation for buyers is significant at $\alpha = 0.001$ ($F_{1,155}$), with an $R^2 = 0.10$. For suppliers, the equation is also significant at $\alpha = 0.001$ ($F_{1,155}$), with an $R^2 = 0.08$. These results do not change even when competitive advantages at time 1 are controlled for the equation above.

The preservation of competitive advantages (Fig. 3) is estimated in a model in which buyers and suppliers are pooled together (i.e., each side’s view of the dyad is treated as an independent observation) and then simultaneously estimated under high/low opportunism (high/low is determined by a median split of the opportunism scale). Differences in the parameters are tested across high/low opportunism and the parameters with significant differences are freely estimated. Model specifics are represented below the figure. In this way, we are able to test and examine the multiple interactions of a complex model structure.

Appendix B. Spill-over effects of competitive advantages

The spill-over effects of competitive advantages are considered via correlations between changes in
competitive advantages over time and changes in profit performance, expectations of continuity, and problem-handling behaviors. Before calculating this, I undertook a descriptive analysis of the longitudinal changes of the summed-scale value\(^5\) for each construct. Table 2 gives the cross-sectional statistics (i.e., time 1 and time 2 measures) and longitudinal statistics (i.e., difference measures) for each construct. Two types of raw difference scores are provided under the longitudinal statistics. The first reflects the average total change for each construct between time 1 and time 2. These numbers are relatively small, with less than a 1-point change over time. While this may suggest that the scores are relatively stable, the second difference score gives greater detail—it summarizes the number and extent of negative and positive changes in the constructs between time 1 and 2. A greater amount of diversity is reflected in these scores, which indicate that 42–56% of the sample changed in a positive or negative direction, and the average magnitude of these changes ranged from 1.1 to 3.5 points. This diversity of changes in each construct across the respondents suggests sufficient variation for meaningful analysis of the interrelationships among change scores.

The most direct measure of change in these constructs is the raw difference \((X_{2} - X_{1})\) in scores for each respondent over time. However, this is an inappropriate measure because the time 1 scores are reflected in the time 2 scores (Lord, 1963); thus, time 2 scores are dependent on the values at time 1, creating (unwanted) variance due to time 1 scores (see Cohen and Cohen, 1983 for more details). To overcome this bias, researchers (Cohen and Cohen, 1983; Cronbach and Furby, 1970)\(^6\) recommend the use of residualized change scores (RCS), which are obtained by regressing the summed-scale value of each construct at time 2 on the summed-scale value for the construct at time 1. This essentially partials out a residual measure that reflects the change from time 1 to time 2 and removes the impact of the initial starting point (time 1). The residuals are the RCS values—these are the amounts of change in the construct at time 2 per unit increase at time 1. The RCS values thus provide comparable measures of change across respondents that are not affected by initial values.

The structure of such regressed or partialled change scores helps clarify the nature of the defect of simple change scores. A simple change score presumes that the regression of time 2 scores on time 1 has a slope of 1. This almost never occurs in the behavioral sciences, as this requires that the correlation between time 1 and time 2 \((r_{12})\) is unity when the standard deviations (SD) are equal, and more generally that \(r_{12} = SD_{2}/SD_{1}\), which is relatively unlikely. Both individual differences and measurement error operate to reduce the \(\beta\) of the partialled score. Thus, the effect of using pure difference scores is typically one of overcorrection of the postscore by the prescore. My interest is in whether changes in competitive advantages are correlated with changes in other key outcome measures. To this end, having estimated the RCS scores, I then estimate the correla-

---

\(^5\) Each respondent’s score to individual items are summed to obtain a summed-scale value of the corresponding construct.

\(^6\) Johnston et al. (1990) are an example of a study in marketing that also employs this technique.
tion between competitive advantages and profit performance, expectations of relationship continuity, and constructive responses to problems using the subset of respondents who completed surveys at both time periods ($N = 321$).

Appendix C. Scale items and reliabilities

$\alpha =$ Cronbach alpha scale reliability.

Likert scales (1 = Strongly Disagree; 7 = Strongly Agree).

Unless otherwise noted, “They” and “us” refer to the two firms, the buyer and supplier together.

All of the scales below, with the exception of interpersonal trust, refer to the organizational relationship between the firms.

C.1. Time 1 measures

Coordination effort ($\alpha = 0.79$)
They work on joint projects tailored to their needs.
They work together to exploit unique opportunities.
Both companies are always looking for synergistic ways to do business together.

Bilateral idiosyncratic investments ($\alpha = 0.76$)
If this relationship were to end, they would be wasting a lot of knowledge that’s tailored to their relationship.
If either company were to switch to a competitive buyer or vendor, they would lose a lot of the investments made in the present relationship.
They have invested a great deal in building up their joint business.

Suspected opportunism of the counterpart ($\alpha = 0.90$)
When a problem occurs, how often will the buyer (supplier) do the following? (1 = Hardly Ever, 7 = Very Often)
They make hollow promises.
They are aloof toward us.
They “window dress” their efforts to improve.
They expect us to pay more than our fair share of the costs to correct the problem.
They are unwilling to accept responsibility.
They make false accusations.
They provide false information.
They fail to provide proper notification.

Goal congruence ($\alpha = 0.87$)
The firms share the same goals in the relationship.
They have compatible goals.
They support each other’s objectives.
They have different goals. (R)

Interpersonal trust * ($\alpha = 0.91$)
Our promises to each other are reliable.
We are very honest in dealing with each other.
We trust each other.
We would go out of our way to help each other out.
We consider each other’s interests when problems arise.
* For this scale only, “Our” and “We” refer to the individual representatives.

C.2. Time 2 measures

Realized competitive advantages ($\alpha = 0.81$)
They have gained strategic advantages over their competitors.
The relationship has not resulted in strategic advantages for them. (R)
They have gained benefits that enable them to compete more effectively in the marketplace.
The relationship has not resulted in strategically important outcomes. (R)

Profit performance ($\alpha = 0.83$)
They have achieved a high level of joint profits between them.
They have generated a lot of profits together.

Constructive responses to problems ($\alpha = 0.71$)
We try to repair problems that occur between us.
We try to work things out when a problem arises.

Expectations of relationship continuity ($\alpha = 0.84$)
Our relationship with this firm will last far in to the future.
We expect to continue working with this firm on a long-term basis.

References


Country of origin: A competitive advantage?

Jagdish Agrawal a, Wagner A. Kamakura b, *

a California State University, Hayward, CA 94542, USA
b Department of Marketing, College of Business Administration, University of Iowa, Iowa City, IA 52242, USA

Received 1 November 1998; accepted 1 October 1999

Abstract

Country of origin has been identified in the literature as an important cue that might be used by global marketers to influence consumers’ valuation of the brand. Its effect on consumer perceptions, affect and behavioral intentions has been widely documented, based on consumer surveys and laboratory experiments. Despite this empirical evidence, we argue that country of origin is only one extrinsic cue among many extrinsic and intrinsic cues available to the consumer in a real purchase situation. Furthermore, in real life, consumers are likely to engage in some level of information search, which would further dilute the country of origin effect in the marketplace. Based on these arguments, we conclude that country of origin might not necessarily lead to a competitive disadvantage in terms of a price premium or discount. For a sample of products, we show that the objective product quality varies significantly by country of origin, and that these differences are consistent with extant research on country of origin effects on consumers’ perceptions. After controlling for quality differences across brands, we demonstrate that marketers from different countries charge prices that are justified by differences in product quality. Price premiums or discounts are therefore explained by differences in product quality rather than the image effect produced by the country of origin cue.

Keywords: Country of origin; Meta-analysis; Hedonic regression; Competitive advantage; Objective quality of brands

1. Introduction

The effect of the country of origin (hereafter referred to as COO) of brands on consumer behavior has been one of the most researched issues in international business (Peterson and Jolibert, 1995). Over the last three decades, several researchers have examined the effect of COO on consumers’ overall evaluation of product quality, beliefs regarding individual attributes of a product, attitude towards brand, and behavioral intention. Several reviews of this body of literature have been published describing the nature and the extent of effects, the circumstances when the effect is more or less pronounced, and the factors moderating the effect (Leifeld, 1993; Peterson and Jolibert, 1995; Samiee, 1994; Verlegh and Steenkamp, 1999). Their conclusions suggest that COO has a significant effect on consumers’ evaluations of products and that consumers tend to use COO as an extrinsic cue to make judgment about the quality of products.

Whether directly through personal experiences, through information acquired from other sources or due to stereotypical beliefs about countries, consumers also tend to develop product–country im-
ages. These are images of quality of specific products marketed by firms associated with different countries (Heslop and Papadopoulos, 1993; Johansson and Thorelli, 1985). A few examples of such product–country images are Colombian coffee, Swiss watches, US appliances, Japanese electronics and German automobiles. Because of the product–country images consumers hold, and their sensitivity to COO, COO is believed to be one way of enhancing brand equity (Keller, 1993; Shocker et al., 1994). If consumers hold a positive (negative) product–country image for a given product and country, this image could lead to a generalized positive (negative) evaluation and attitude towards all the brands of a product associated with that country. Such COO-based equity might even extend to other product categories due to stereotypical bias.

One of the consequences of brand equity is that the brand may command premium prices (Aaker, 1996; Keller, 1993). Given the empirical evidence in support of consumers’ judgment of quality based on the COO, one would expect such COO effects to influence the firms’ pricing decisions. More specifically, firms originating in countries with better product–country image should be able to charge premium prices and those firms originating in countries with poor product–country image may have to offer products at discounted prices due to their country image. This price premium/discount should exist above and beyond the price differential due to quality differences. If the COO is found to affect pricing decisions, then COO can be regarded as a major variable leading to competitive (dis)advantages to firms originating in certain countries.

1.1. Role of COO on the consumers’ decision-making process

Recently, Peterson and Jolibert (1995) and Verlegh and Steenkamp (1999) have conducted comprehensive meta-analyses of the literature on COO effects. These researchers have examined the relative impact of COO on different stages of the consumers’ decision-making process such as perception, attitude and behavioral intention. They have also examined the moderating effect of several study characteristics. The findings of these two studies provide some important insights regarding COO’s possible effect in real markets.

An important conclusion of these two studies is that the effect of COO is smaller for multi-cue studies than for single cue studies. One of the criticisms of COO studies is that many COO studies have manipulated only one cue, i.e., COO cue. When all the other information are controlled, COO is likely to have a significant impact on product evaluation. However, in a real purchasing situation, consumers are likely to have additional information and access to other cues such as the actual physical product, brand name, price, warranty, etc. In such a situation, the impact of any one single cue such as COO may diminish significantly. Peterson and Jolibert (1995) report a significant decrease in the effect of COO on both quality perception and purchase intention when multiple cues are considered compared to COO as the only cue. Similarly, Verlegh and Steenkamp (1999) also report a significant decrease in the effect of COO in multiple cue studies compared to single cue studies. Thus, the quantitative reviews of the empirical results of previous studies clearly show that COO’s effect reduces significantly in the presence of other cues.

Another important result of these reviews is regarding the role of COO on influencing different stages of consumer behavior. Results of these reviews clearly suggest that although COO plays an important role in product evaluation, the effect tends to become weaker as one moves from perception of product quality to attitude formation and to behavioral intention. Peterson and Jolibert (1995) reported a significant decrease in COO effect as one moved from quality perception in single cue studies to purchase intention in single or multiple cue studies. Similarly, Verlegh and Steenkamp (1999) also reported significantly larger effect for quality perception compared to attitude formation and purchase intention. In other words, COO has significantly lesser impact as consumers move closer to the actual purchase situation from belief formation regarding the relative quality of brands.

Taken together, the empirical evidence from previous studies on COO effects show that as consumers move closer to actual choice behavior, COO effect tends to become weaker. Compared to situa-
tions in which COO effects have been examined so far, the actual purchase decisions consumers make in their daily lives carry greater potential risks and benefits. A consumer must not only incur the costs associated with the purchase but also live with the consequences of his/her choice decisions. Therefore, it is quite reasonable to expect the consumer to be willing to allocate more processing effort in a real life decision than in a hypothetical scenario considered in most previous studies of COO effect. Moreover, in the real consumer decision-making environment, COO as an informational cue competes head-to-head with other extrinsic cues and intrinsic cues. Consequently, the relative effect of COO on actual choice behavior is likely to be small. Therefore, if consumers do not use COO as an important informational input in the actual purchase of products, COO is unlikely to influence pricing decisions of firms.

In their attempt to conserve cognitive capacity, consumers are also known to switch to simplifying heuristics as the decision-making environment becomes more complex (Bettman, 1979). In such situations, extrinsic cues can be used to represent information regarding the known quality of a brand (Han, 1989). In other words, consumers might use extrinsic cues such as COO as a summary construct representing their knowledge about brands from different countries (Han, 1989). For example, consumers might rate microwave ovens originating from a certain country higher than others, not because they infer product quality from the COO cue, but because they know that microwave ovens originating from that particular country do indeed have better quality. In this situation, COO is used to eliminate brands and develop an evoked set rather than make inferences about quality, thus, saving consumers from extensive evaluation of intrinsic attributes. Thus, as indicated by previous research, COO may still influence attitude because it reflects the consumers’ knowledge about product quality. However, because one of the necessary conditions for premium price is the asymmetry in information between buyers and sellers (Rao and Monroe, 1989), if consumers are already knowledgeable about product quality (to the point of using COO as a summary of their knowledge), they are not likely to pay any price that is not justified by that quality.

Therefore, there are two main reasons to expect weaker COO effects in the marketplace. First, consumers are likely to invest more of their cognitive resources when making purchase decisions in real life where COO is only one of the many cues to be processed, and as a consequence, COO effects are likely to be weaker. Second, consumers might use COO not as a cue to infer quality, but as a way of summarizing their knowledge about quality in the particular product category they are choosing from. Because of their greater involvement with the actual decision-making task, and their use of COO as a summary of their knowledge about product quality, consumers are not likely to pay prices that are not justified by product quality.

These results tend to suggest that although COO may make a significant influence on the consumers’ judgment of the quality of products in laboratory experiments, its role in influencing consumers’ actual choice behavior may be quite limited in the presence of other information and cues. The COO is simply one of the several cues available to consumers. As consumers move along in their decision-making process from assessment of quality of brands to attitude formation and final choice decision, the role of any one single cue such as COO may be quite insignificant. Besides, other factors such as budget constraint and need urgency may further moderate the actual choice behavior. In such a circumstance, consumers may not be willing to pay premium price or expect discounts simply because of the COO of brands. Even though COO might play an important role in affecting consumer perceptions of brand quality in a laboratory experiment, this effect may not necessarily lead to price premium or discounts in the marketplace.

Although there are many examples of manufacturers emphasizing country image in their promotional campaigns, packaging or branding decisions (see Leclerc et al., 1994; Papadopoulos, 1993), very little is known regarding the influence of COO on pricing decisions. In other words, although previous research has provided strong evidence regarding the effect of COO on consumer behavior, little is known about how this observed effect has influenced firms’ behavior. Given the ample empirical evidence for the effect of COO on consumer behavior, there is a need for a systematic research on the market value gener-
ated by the COO cue, and how this cue may have benefited firms associated with countries possessing positive product–country images (Leifeld, 1993; Samiee, 1994).

1.2. Objectives of the study

Two objectives have guided this research study. First, we examine differences in the level of objective quality at the COO level using longitudinal data on quality of competing brands associated with different countries in a sample of product categories. Our focus is on country of origin (COO) as opposed to country of manufacture (COM). COO refers to the country with which the firm producing a brand is associated, whereas COM refers to the country where a brand is actually manufactured or assembled. Previous studies have focused on assessing consumers’ perception of quality of brands associated with different countries. They have not examined the source of these perceptual differences. Their findings raise the question whether observed perceptual differences in product quality associated with different countries are due to some halo effect indicating perceptual bias in favor of products made in some countries or due to actual differences in objective quality across firms representing different countries. In addition, although researchers have examined whether or not market prices of products reflect their relative quality (Tellis and Wernerfelt, 1987), we are not aware of any systematic study to examine whether there is a significant difference in the actual quality of products associated with different countries. A correspondence between perceptions of differences in quality of brands reported in previous studies and differences in objective quality of brands across countries might reflect consumers’ knowledge of the actual quality of the products marketed by different countries, rather than their perceptual bias. Moreover, a significant difference in quality of the products across countries may also suggest that COO could be useful as a cue to make judgments about product quality.

The second objective of this study is to assess whether or not COO effects on consumer behavior have resulted in firms charging premiums/discounts on their products on the basis of their product–country images. Despite a large number of experiments and surveys measuring the effects of COO on perception and affect, we are not aware of any major attempt to measure the impact of COO on the firms’ behavior. Therefore, unlike previous studies of COO effects on consumers’ judgment of quality and behavioral intentions, this study is designed to assess the impact of COO on managerial decision. More specifically, we empirically test whether manufacturers charge price premiums or price discounts attributable to COO, after accounting for the differences in objective quality of brands. Previous studies have provided ample evidence that consumers associate COO with their perception of product quality. As a natural extension of this research stream, this study seeks to know whether consumers also end up paying premium (or discount) prices for the brands associated with certain countries due to their better (or poor) product quality images.

In the following sections, we first present the description of data from Consumer Reports used in this study. Next, we examine the differences in the objective quality of brands associated with different countries for a sample of products. In order to test for COO-based (dis)advantage in the marketplace, we then use a series of hedonic price regressions estimating price premiums (or discounts) for brands associated with different countries after accounting for their quality differences. Through a meta-analysis of the hedonic regression results, we test for the existence of price premiums or discounts attributed to COO of brands. Section 4 presents a summary and discussion of results.

2. COO and objective product quality

2.1. Data

Data on objective quality were collected for a sample of 13 products from various issues of the US consumer magazine Consumer Reports from 1980 through 1994 (Consumer Union of US, 1980–1994). Details about the data are presented in Table 1. Some of the products sampled for this study have been used in previous studies on COO effects. The products were selected such that firms from more than one country were competing in that product category. All products (except for video tapes) included are durable goods, and most of them are in the...
Table 1
Summary of the data used in each hedonic price regression

<table>
<thead>
<tr>
<th>Study</th>
<th>Product</th>
<th>Date</th>
<th>Brands</th>
<th>Countries</th>
<th>Features</th>
<th>Price (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Telephone</td>
<td>1/89</td>
<td>25</td>
<td>2</td>
<td>9</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>Telephone</td>
<td>5/86</td>
<td>15</td>
<td>2</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>Telephone</td>
<td>12/92</td>
<td>28</td>
<td>2</td>
<td>10</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>Answering machine</td>
<td>5/83</td>
<td>17</td>
<td>2</td>
<td>8</td>
<td>137</td>
</tr>
<tr>
<td>5</td>
<td>Answering machine</td>
<td>5/86</td>
<td>24</td>
<td>2</td>
<td>13</td>
<td>147</td>
</tr>
<tr>
<td>6</td>
<td>Answering machine</td>
<td>11/91</td>
<td>18</td>
<td>2</td>
<td>19</td>
<td>112</td>
</tr>
<tr>
<td>7</td>
<td>Camcorder</td>
<td>3/94</td>
<td>12</td>
<td>4</td>
<td>9</td>
<td>738</td>
</tr>
<tr>
<td>8</td>
<td>Camcorder</td>
<td>11/87</td>
<td>25</td>
<td>3</td>
<td>10</td>
<td>1051</td>
</tr>
<tr>
<td>9</td>
<td>Camera</td>
<td>8/83</td>
<td>26</td>
<td>3</td>
<td>15</td>
<td>94</td>
</tr>
<tr>
<td>10</td>
<td>Camera</td>
<td>9/86</td>
<td>27</td>
<td>2</td>
<td>11</td>
<td>111</td>
</tr>
<tr>
<td>11</td>
<td>Camera</td>
<td>11/88</td>
<td>29</td>
<td>2</td>
<td>14</td>
<td>146</td>
</tr>
<tr>
<td>12</td>
<td>Camera</td>
<td>12/92</td>
<td>26</td>
<td>4</td>
<td>18</td>
<td>207</td>
</tr>
<tr>
<td>13</td>
<td>CD player</td>
<td>3/90</td>
<td>28</td>
<td>3</td>
<td>11</td>
<td>322</td>
</tr>
<tr>
<td>14</td>
<td>CD player</td>
<td>3/91</td>
<td>24</td>
<td>3</td>
<td>11</td>
<td>330</td>
</tr>
<tr>
<td>15</td>
<td>CD player</td>
<td>3/92</td>
<td>21</td>
<td>3</td>
<td>10</td>
<td>264</td>
</tr>
<tr>
<td>16</td>
<td>CD player</td>
<td>3/93</td>
<td>26</td>
<td>3</td>
<td>9</td>
<td>300</td>
</tr>
<tr>
<td>17</td>
<td>CD player</td>
<td>3/94</td>
<td>28</td>
<td>4</td>
<td>11</td>
<td>207</td>
</tr>
<tr>
<td>18</td>
<td>CD player</td>
<td>5/87</td>
<td>24</td>
<td>3</td>
<td>13</td>
<td>282</td>
</tr>
<tr>
<td>19</td>
<td>CD player</td>
<td>6/85</td>
<td>14</td>
<td>3</td>
<td>12</td>
<td>468</td>
</tr>
<tr>
<td>20</td>
<td>Clock radio</td>
<td>9/86</td>
<td>19</td>
<td>3</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>21</td>
<td>Clock radio</td>
<td>9/89</td>
<td>22</td>
<td>3</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>22</td>
<td>Clock radio</td>
<td>11/80</td>
<td>22</td>
<td>3</td>
<td>14</td>
<td>53</td>
</tr>
<tr>
<td>23</td>
<td>Clock radio</td>
<td>11/92</td>
<td>18</td>
<td>3</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>24</td>
<td>Cordless phone</td>
<td>9/83</td>
<td>12</td>
<td>2</td>
<td>7</td>
<td>204</td>
</tr>
<tr>
<td>25</td>
<td>Cordless phone</td>
<td>11/86</td>
<td>16</td>
<td>2</td>
<td>10</td>
<td>146</td>
</tr>
<tr>
<td>26</td>
<td>Cordless phone</td>
<td>11/89</td>
<td>17</td>
<td>2</td>
<td>11</td>
<td>149</td>
</tr>
<tr>
<td>27</td>
<td>Cordless phone</td>
<td>11/91</td>
<td>18</td>
<td>2</td>
<td>13</td>
<td>113</td>
</tr>
<tr>
<td>28</td>
<td>Microwave</td>
<td>3/81</td>
<td>21</td>
<td>4</td>
<td>8</td>
<td>508</td>
</tr>
<tr>
<td>29</td>
<td>Microwave</td>
<td>3/89</td>
<td>21</td>
<td>4</td>
<td>15</td>
<td>215</td>
</tr>
<tr>
<td>30</td>
<td>Microwave</td>
<td>5/83</td>
<td>16</td>
<td>3</td>
<td>8</td>
<td>385</td>
</tr>
<tr>
<td>31</td>
<td>Microwave</td>
<td>6/94</td>
<td>20</td>
<td>4</td>
<td>10</td>
<td>199</td>
</tr>
<tr>
<td>32</td>
<td>Microwave</td>
<td>11/85</td>
<td>17</td>
<td>3</td>
<td>8</td>
<td>278</td>
</tr>
<tr>
<td>33</td>
<td>Microwave</td>
<td>11/86</td>
<td>20</td>
<td>4</td>
<td>9</td>
<td>229</td>
</tr>
<tr>
<td>34</td>
<td>Microwave</td>
<td>11/91</td>
<td>17</td>
<td>4</td>
<td>15</td>
<td>220</td>
</tr>
<tr>
<td>35</td>
<td>Microwave</td>
<td>12/92</td>
<td>20</td>
<td>4</td>
<td>18</td>
<td>154</td>
</tr>
<tr>
<td>36</td>
<td>Receiver</td>
<td>3/89</td>
<td>25</td>
<td>2</td>
<td>16</td>
<td>230</td>
</tr>
<tr>
<td>37</td>
<td>Receiver</td>
<td>3/93</td>
<td>29</td>
<td>3</td>
<td>15</td>
<td>335</td>
</tr>
<tr>
<td>38</td>
<td>Receiver</td>
<td>3/94</td>
<td>25</td>
<td>2</td>
<td>11</td>
<td>313</td>
</tr>
<tr>
<td>39</td>
<td>Receiver</td>
<td>7/88</td>
<td>26</td>
<td>2</td>
<td>17</td>
<td>286</td>
</tr>
<tr>
<td>40</td>
<td>TV set</td>
<td>1/81</td>
<td>18</td>
<td>2</td>
<td>13</td>
<td>665</td>
</tr>
<tr>
<td>41</td>
<td>TV set</td>
<td>3/87</td>
<td>39</td>
<td>4</td>
<td>20</td>
<td>461</td>
</tr>
<tr>
<td>42</td>
<td>TV set</td>
<td>3/94</td>
<td>22</td>
<td>5</td>
<td>13</td>
<td>543</td>
</tr>
<tr>
<td>43</td>
<td>TV set</td>
<td>5/88</td>
<td>24</td>
<td>5</td>
<td>20</td>
<td>285</td>
</tr>
<tr>
<td>44</td>
<td>VCR</td>
<td>1/87</td>
<td>19</td>
<td>3</td>
<td>15</td>
<td>814</td>
</tr>
<tr>
<td>45</td>
<td>VCR</td>
<td>3/93</td>
<td>23</td>
<td>5</td>
<td>16</td>
<td>341</td>
</tr>
<tr>
<td>46</td>
<td>VCR</td>
<td>3/94</td>
<td>12</td>
<td>5</td>
<td>10</td>
<td>254</td>
</tr>
<tr>
<td>47</td>
<td>VHS tape</td>
<td>9/88</td>
<td>51</td>
<td>4</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>48</td>
<td>VHS tape</td>
<td>9/90</td>
<td>14</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>49</td>
<td>VHS tape</td>
<td>11/84</td>
<td>21</td>
<td>4</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>50</td>
<td>VHS tape</td>
<td>11/86</td>
<td>35</td>
<td>4</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>
consumer-electronics category. After excluding brands whose COO could not be identified, the total sample represented 122 brands originating in seven countries (US, Japan, Korea, Netherlands, Sweden, France and Germany). Because of the limited sample size, brands from Sweden, France and Germany are grouped into the “Others” category.

2.2. Objective quality and COO

In order to measure the differences in the level of objective quality of brands across countries, we compared the quality rankings of brands for each product within a given Consumer Reports test. This provided us with data on the number of times each country dominated other countries in quality of brands of a product at a given time. This procedure was repeated for each product and for each Consumer Reports test. Finally, we consolidated the results of these comparisons to obtain a summary of all pair-wise comparisons of brands across time (1980–1994) and products. Results of this comparison are presented in Table 2. This table shows the number of pair-wise comparisons in which a country dominated another country in brand quality in the sampled product categories. For example, there was a total of 4734 (2806 + 1928) direct comparisons between brands originating in Japan and the US. In 2806 (59%) of these comparisons, the quality of the Japanese brands was ranked higher than that of the US brands, whereas in the rest (41% of the cases), the quality of the US brands was ranked higher than that of the Japanese brands.

Table 2
Paired comparisons of overall quality (column countries dominate row countries)

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Japan</th>
<th>Korea</th>
<th>Netherlands</th>
<th>Others*</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>0</td>
<td>2806</td>
<td>110</td>
<td>241</td>
<td>200</td>
</tr>
<tr>
<td>Japan</td>
<td>792</td>
<td>0</td>
<td>108</td>
<td>378</td>
<td>284</td>
</tr>
<tr>
<td>Korea</td>
<td>183</td>
<td>420</td>
<td>16</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>Others</td>
<td>232</td>
<td>353</td>
<td>34</td>
<td>31</td>
<td>0</td>
</tr>
</tbody>
</table>

*Others include Germany, France and Sweden.

In 2806 of the (2806 + 1928) comparisons of brands associated with Japan and US, Japanese brands were superior in rank to US brands.

We applied the logistic equivalent of Thurstone’s Case V (i.e., using a logistic function instead of a normal ogive to translate proportions into scale values) to the data in Table 2 in order to obtain maximum-likelihood estimates and standard errors of overall quality scores for each country. The estimates, along with their 95% confidence intervals using the US as the standard (i.e., US = 0), are displayed in Fig. 1.

The estimates displayed in Fig. 1 clearly show that for the sampled products, the overall objective quality (as measured by Consumer Reports) of products originating in Japan is significantly better than the products originating in the US, Europe and Korea — the 95% confidence interval for Japan is clearly above zero (the standard level of quality set for the US), and higher than the confidence intervals for Korea and Others. On the other hand, the quality of Korean products is significantly lower than that of the US, Japanese and European products. In summary, these results show that countries vary significantly in terms of the objective quality of the products they market in the US.

2.3. Correspondence between objective and perceived qualities

A number of studies have examined consumers’ judgment of quality of durable products including some of the products included in this study. Results of these studies are presented in Table 3. The table includes results from only those studies where consumers in the US judged quality of products made in the countries included in this study, in order to be...
Table 3
Summary of studies showing average quality rating of durables of different countries

<table>
<thead>
<tr>
<th>Authors</th>
<th>Products</th>
<th>Variable</th>
<th>Japan</th>
<th>USA</th>
<th>Germany</th>
<th>Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chao (1989)</td>
<td>TV</td>
<td>Quality</td>
<td>–</td>
<td>4.79</td>
<td>–</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>VCR</td>
<td>6 = excellent,</td>
<td>–</td>
<td>4.51</td>
<td>–</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td>Stereo</td>
<td>1 = poor</td>
<td>–</td>
<td>4.62</td>
<td>–</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 = high, 1 = low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Cordell (1992)</td>
<td>Wrist watch</td>
<td>Quality</td>
<td>0.46</td>
<td>–</td>
<td>0.56</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = same as Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = better, – 1 = worse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Elliot and Cameron (1994)</td>
<td>Dishwasher</td>
<td>Quality</td>
<td>–</td>
<td>–</td>
<td>4.70</td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 = very good, 1 = very poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Han and Terpstra (1988); Han (1989)</td>
<td>TV</td>
<td>Overall image</td>
<td>6.01</td>
<td>5.24</td>
<td>4.68</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 = high, 1 = low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Han and Terpstra (1988); Han (1989)</td>
<td>Auto</td>
<td>Overall image</td>
<td>5.75</td>
<td>4.55</td>
<td>5.56</td>
<td>3.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 = high, 1 = low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 = very favorable,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = least favorable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Johansson et al. (1994)</td>
<td>Tractor</td>
<td>Quality</td>
<td>4.90</td>
<td>6.10</td>
<td>5.60</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 = very good, 1 = very poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Levin et al. (1993)</td>
<td>Auto</td>
<td>Quality</td>
<td>1.38</td>
<td>0.82</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 4 = highest score,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>− 3 = lowest score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Maheswaran (1994)</td>
<td>PC</td>
<td>Favorableness</td>
<td>5.97</td>
<td>–</td>
<td>–</td>
<td>4.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Higher scores better</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Tse and Lee (1993)</td>
<td>Stereo system</td>
<td>Performance</td>
<td>4.43</td>
<td>–</td>
<td>–</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 = very good, 1 = very bad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Wall and Heslop (1986)</td>
<td>Home</td>
<td>Quality</td>
<td>6.30</td>
<td>5.50</td>
<td>5.20</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>entertainment</td>
<td>7 = very good, 1 = very poor</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>
</tbody>
</table>

comparable to the objective measure of relative quality displayed in Fig. 1. Pair-wise comparisons of average scores of perceived quality of products across countries show that in all cases, the quality of products from Japan, the USA and Germany are perceived to be better than products originating from South Korea. Furthermore, in about 80% of the cases, the Japanese products are perceived to be better than American and German products; and American products are perceived to be better than German products in 75% of the cases. In summary, considering all the comparisons of mean perceived quality of products originating from different countries, consumers tend to perceive Japanese products to be better than the rest of the countries studied; American products are perceived to be better than German and Korean products; and German products are perceived to be better than Korean products.
In spite of the differences in the measurement of quality between this study and other studies, there are some important similarities in results. A cursory comparison of our measurement of objective quality (Fig. 1) and perceived quality of products from different countries as reported in other studies (Table 3) indicates that consumers’ quality assessments are consistent with the observed differences in product quality across these countries. This correspondence between perceived and objective quality indicates that consumers’ assessment of quality for products originating from different countries might be based on factual information such as their own experiences with these products or information obtained from neutral sources such as buying guides. Furthermore, the differences in quality perceptions observed in previous COO studies could be a reflection of the consumers’ knowledge of the products originating from different countries, rather than a perceptual bias.

3. COO and price premium/discount

The second objective of this study is to examine whether firms associated with different countries charge a premium price or give a price discount for their brands that is above and beyond what would be justified by their quality differences. We used hedonic price regression to estimate the price premiums above and beyond what would be charged for the same product characteristics. The major question we want to address is: are producers from different countries charging price premiums or offering price discounts because of advantages or disadvantages from their overall COO image, but not justified by their product features?

Following the usual practice in hedonic price analysis, we fit the following function to market data, through regression analysis,

$$\ln P_j = \alpha_0 + \sum_{i=2}^{f} \alpha_i D_{ij} + \beta Z_j + e_j,$$

where $j$ denotes a brand; $P_j$ is the street price for brand $j$; $i$ denotes a country, $D_{ij}$ is equal to 1 if brand $j$’s COO is country $i$; zero otherwise, and fixed to zero for the US as the standard of comparison ($i = 1$); $Z_j$ is the vector of features for brand $j$; $e_i$ is the price surplus - (relative to US domestic products) (if $e_i > 1$) or discount (if $e_i < 1$) for brands originating in country $i$; $\beta$ is a vector -of parameters adjusting prices for differences in objective quality across brands.

The intercept ($\alpha_0$) in the hedonic regression above captures the average price for US domestic products.
while the dummy coefficients ($\alpha_r$) measure the log-price differentials (relative to US products) due to COO.

The hedonic price regression in Eq. (1) was applied to the data published by Consumer Reports for 12 products in the 1980–1994 period, in 50 separate studies. One regression was estimated across all brands within a given product category being reviewed by Consumer Reports in a given issue. The application of the hedonic regression model presented in Eq. (1) to the data summarized in Table 1 led to estimates of the price premium/discount ($e^{\alpha_r}$), above/below the expected price from the products’ measured quality, which can be attributed to the products’ COO. Because product quality is evaluated along attributes that are specific to each product category, the COO price premium/discounts are estimated within each particular study published by Consumer Reports. We treat each of these hedonic regressions as an independent study, leading to independent estimates of COO effects (i.e., price premiums). These estimates of the price premium/discount relative to products originating in the US, across all 50 studies are summarized in the box-and-whiskers plot in Fig. 2, which shows the median, bottom and top quartiles within each country. The effects of the other predictors in the hedonic regressions are not reported for two main reasons: (a) they are not of immediate interest for this particular study, and (b) they cannot be easily summarized, because product attributes are not consistent across product categories and time.

Notice, from Fig. 2, the presence of four clear outliers among all the estimates, for cameras and microwave ovens from other European countries, microwaves ovens from Japan, and Dutch radio clocks. These outliers were excluded from subsequent analyses. Fig. 2 also suggests that there is considerable discrepancy in COO effects across the 50 studies, with only minor differences in the median effects for each country. Products originating in Japan seem slightly more expensive (after accounting for objective quality) than the US products, while those originating in The Netherlands and Korea are discounted slightly, relative to American products.

3.2. A meta-analysis of COO effects on price

Because the estimates summarized in Fig. 2 were obtained from 50 independent studies of a dozen different product categories, a substantial portion of the variance in COO effects across studies might have arisen due to differences in design across studies. Therefore, we test for the existence of COO-related price premiums/discounts through a meta-analysis of these 50 studies, in an attempt to better generalize from their results. Meta-analysis has been used in many disciplines as a methodology to draw generalizations from the results obtained in a large number of independent studies (see Farley and Lehman 1986 for a comprehensive discussion of this methodology). Recently, this methodology has also been applied to explore the generalizability of studies on COO effects (Peterson and Jolibert, 1995; Verlegh and Steenkamp, 1999).

Similar to other marketing applications, our main objective is to obtain a better measurement of the phenomenon under study, after sorting out the effects of methodological factors in each of the independent studies. In our meta-analysis of the 50 hedonic price regressions, we considered two discrete (COO and Product category) and four continuous (Number of brands, Number of attributes, Number of countries and Average price) design characteristics. We also included all first-order interactions involving the

---

1 The number of brands for each country included in hedonic regressions were USA (49), Japan (41), Netherlands (5), South Korea (3), Sweden (3), Germany (2), and France (2).
Table 4
ANOVA results for the meta-analysis of price premiums/discounts

<table>
<thead>
<tr>
<th>Source</th>
<th>Weighted sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
<th>F-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>COO</td>
<td>2.66</td>
<td>3</td>
<td>0.89</td>
<td>2.20</td>
<td>0.09</td>
</tr>
<tr>
<td>Product category</td>
<td>5.42</td>
<td>11</td>
<td>0.49</td>
<td>1.22</td>
<td>0.29</td>
</tr>
<tr>
<td>Number of attributes</td>
<td>0.76</td>
<td>1</td>
<td>0.76</td>
<td>1.89</td>
<td>0.17</td>
</tr>
<tr>
<td>Number of countries</td>
<td>0.03</td>
<td>1</td>
<td>0.03</td>
<td>0.06</td>
<td>0.80</td>
</tr>
<tr>
<td>Number of brands</td>
<td>0.38</td>
<td>1</td>
<td>0.38</td>
<td>0.93</td>
<td>0.34</td>
</tr>
<tr>
<td>Average price</td>
<td>0.10</td>
<td>1</td>
<td>0.10</td>
<td>0.24</td>
<td>0.63</td>
</tr>
<tr>
<td>Residual</td>
<td>36.05</td>
<td>89</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (corrected)</td>
<td>44.50</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2 = 0.19$

COO factor, with the exception of the COO versus Product category interaction, because of the co-linearity between these two design characteristics. The main effects allow us to verify whether the estimate of COO effect was influenced by any of the characteristics of the hedonic price regressions. The first-order interactions allow us to verify whether certain design characteristics could have distorted the estimate of the price premium/discount for any particular country. Our dependent variable is the estimate of COO effect ($\alpha$) for each country in each hedonic price regression.

Meta-analyses are typically performed with ANOVA or linear regression, linking the estimates of the phenomenon under study to the variables describing the study design. We use a linear regression in our study. However, because COO effects displayed in Fig. 2 are estimates obtained from various hedonic regressions, the dependent variable in our meta-analysis has an estimation error associated with it. Therefore, we perform our meta-analysis through weighted least squares, using a weight that is inversely proportional to the standard error of each estimate of price premium/discount.

Fig. 3. Mean estimates (and 95% confidence intervals) of price premium/discount by COO.
Results from our meta-analysis, displayed in Table 4, show that even after accounting for differences in study design, one cannot find any generalizable COO effect \((p = 0.09)\) across the 50 hedonic price regressions. A comparison between the results in Table 4 with those from an analysis of variance including interactions between the price premium/discount and the design characteristics showed no significant interactions, leading to the conclusion that the estimates of COO price premium/discount in these studies were not affected by any of the design characteristics (at the \(p = 0.05\) level). The 95% confidence intervals for the mean COO effect in Fig. 3 show that none of the four countries considered had a significant price premium/discount relative to the US.

This analysis indicates that after considering the price differences in brands due to quality differences, COO has no significant influence on prices. In other words, consumers do not seem to pay more or less because they hold better or worse image regarding the quality of products originating in different countries. There is no price premium or discount for brands originating in different countries, once their quality differences are accounted for.

4. Summary and conclusions

Because of the growing competition from international firms, consumers’ sensitivity to COO has become a relevant issue for brand managers. Researchers have studied this issue extensively by examining consumers’ judgments of quality, their preferences and behavioral intentions in response to the COO cue. However, despite the large number of experiments and surveys measuring the effects of COO on perception and affect, we are not aware of any major attempt to measure the impact of COO on actual markets. The measurement of COO effect on actual markets is essential because effects observed in controlled experiments may not necessarily hold in real markets. In the particular case of COO effects, there are at least two main reasons why these effects might not hold in a real, competitive market. First, when making actual purchase decisions for many of the products tested in our and other studies, consumers are likely to devote more time and processing capacity than in laboratory studies involving remote and hypothetical situations. They are also likely to engage in some level of information search, so that COO will be only one cue competing with many other sources of information. Second, when asked for their perceptions and attitudes towards products from various countries, consumers might express their knowledge about the actual quality of those products, using COO not as a non-substantive cue of quality, but as a summary of their knowledge.

In a competitive market, consumers will be willing to pay only prices that reflect their perception of quality for the available products. If consumers’ perceptions are consistent with the actual quality, manufacturers will be pressed to charge only prices that are justified by that quality, and therefore, would not be able to charge a premium, nor will be motivated to offer a discount.

Integrative reviews of previous studies on consumers’ reactions to COO of brands show that consumers tend to infer brand quality from COO as a cue (Peterson and Jolibert, 1995; Verlegh and Steenkamp, 1999). But the question still remains: is COO a valid indicator of objective quality? A significant difference in quality of brands originating in different countries reported in this study indicates that COO is a reliable cue of brand quality, at least for the products sampled in this study. Although there are several studies that have examined the relationship between price and objective quality (see Ratchford et al. (1996) for a review), we are not familiar with any study that has examined whether COO is a valid indicator of objective quality or not. Our results, based on longitudinal data for a sample of products, suggest that for certain products, COO might be a valid cue for relative product quality.

Our meta-analysis of results from a series of hedonic regression analyses of price on physical attributes and COO suggests that COO does not necessarily result in price premiums or price discounts beyond what would be expected on the basis of objective quality. Although our analyses revealed that countries differed significantly regarding the quality of their products, the hedonic price analyses showed that, after accounting for the differences in quality (performance and features), COO had no significant impact on price.
The products analyzed in this study are such that intrinsic information about product quality is readily available to consumers through buying guides such as Consumer Reports and other sources such as advertising, retailers, etc. These products are "utilitarian" or "search" goods. Consumers, based on personal experience or information from other sources, appear to have developed knowledge regarding the quality of products made in different countries, and might use COO as a summary construct rather than as an inferential cue to make judgments about the quality of brands. Therefore, one explanation for the lack of premium price for Japanese products is that, in the face of the consumers' prior knowledge and availability of objective information, Japanese firms could not charge premium price.

However, for image or "hedonic" products such as wines and fragrances, quality cannot be assessed prior to purchase. In such a situation, extrinsic cues such as price, brand name and COO may be utilized to make judgments about quality (Steenkamp, 1990). It is quite possible that pricing strategies would be different for these "hedonic" products, leading to price premiums and discounts above and beyond quality differences. Unfortunately, factual information regarding intrinsic attributes and objective quality of brands for such products is rarely available. Due to the experiential nature of these products, attribute information is usually based on the perceptions of a sample of consumers or experts, which are often tainted by the known or assumed COO of the products. This paucity of information on objective quality makes it difficult to detect the presence of price premium or discounts over and above the prices justified by product quality in the marketplace.

Another reason for not observing a price premium might be the specific pricing strategies followed by firms from certain countries. The underlying assumption in hedonic regression analysis is that manufacturers set prices to maximize profits in the short/medium run. However, Japanese firms are known to follow a long-term strategy focused on gaining high market shares (Grossberg, 1990) and to price their products accordingly. In this situation, in spite of the fact that a positive product–country image may allow Japanese firms to charge premium prices, they might not do so in order to gain higher market shares. We could not test this possibility because we had no access to market share information for the same products we had used in our study. Future research may examine this issue for products where market share data are available.

Still another reason for the lack of price premium for Japanese products may be competition. Not only do Japanese companies have to compete with products marketed by American, European, Korean and others, but they also have to compete with each other. Therefore, even though Japanese products may enjoy brand equity due to their superior quality, intra-competition among Japanese firms may discourage them to take advantage of this brand equity.

Recent meta-analyses of studies on COO effects on quality perception, attitude formation and behavioral intentions also provide some insights regarding the possible reasons for not observing effect on pricing decisions (Peterson and Jolibert, 1995; Verlegh and Steenkamp, 1999). The findings from these meta-analyses of the results of previous studies on COO’s influence on consumer behavior suggest that COO effect on choice behavior is less pronounced than the effect on perception of quality. Given this lack of strong effect of COO on final choice behavior, COO is unlikely to play any significant role in pricing decisions. The absence of a significant COO effect on price after controlling for the quality difference provides some support to the argument that COO has not influenced the firms’ pricing decisions significantly.

Acknowledgements

This study was partially funded by a grant from the Institute for Industrial Competitiveness, University of Pittsburgh, Pittsburgh.

References


