Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion

Shahar Ayal\textsuperscript{a}, Jérémy Celse\textsuperscript{b,*}, Guy Hochman\textsuperscript{a}

\textsuperscript{a} Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, Israel
\textsuperscript{b} Burgundy School of Business, CEREN EA7477, Université de Bourgogne Franche-Comté, Dijon, France

**ARTICLE INFO**

Keywords:
Fare evasion
Field experiment
Unethical behavior
Public transportation
Behavioral ethics

Jet codes:
C93
K42
B41

**ABSTRACT**

The impact of watching eye cues and descriptive social norm messages on fare evasion was studied in two experiments that were conducted in two railway stations in France. In Study 1, a natural field experiment, passengers were exposed for a two-week period to either a control eye-cues poster or to an experimental eye-cues with a social norm messaging campaign. In Study 2, an artefactual experiment in the field, participants in the experimental train station were asked to participate in a lying task before and after they were exposed to the messaging campaign. The results from both studies suggest that although watching eye cues alone are not effective in a crowded train station, exposing passengers to watching eye cues together with a descriptive social norm messaging campaign reduced the fare evasion rates observed by standard inspection operations, and eliminated lying behavior measured by the die-under-cup paradigm. These results strengthen the external validity of laboratory experiments that have documented small scale cheating behavior and highlight the potential benefits of internal enforcement techniques to fight dishonesty in the field. In addition, they stress the advantage of combining visibility cues and social norms when orienting people toward more moral behavior.

1. Introduction

Small scale cheating, fraud, racketeering, and other forms of unethical behavior on the part of ordinary people are among the greatest personal and societal challenges of our time (Ariely, 2012; Bazerman & Tebrunsel, 2012). On the individual level, the misdeeds of everyday people appear relatively small and isolated. However, because so many people are engaged in these misbehaviors, when considered cumulatively, they cause considerable societal damage and add up to millions of dollars (Ariely, 2012; Barkan, Ayal, & Ariely, 2015). For instance, tax evasion in the US is estimated to be close to $500 billion a year (Cebula & Feige, 2012) and roughly €80 billion a year in France.\textsuperscript{1} Insurance fraud costs over $80 billion annually, and these costs are passed on to customers (Warren & Schweitzer, in press). In addition to their direct costs, research in behavioral ethics shows that dishonesty on the part of ordinary people chips away trust, encourage negative social norms, and increases the prevalence and spread of other unethical behaviors (Kirchler, Hoetlz, & Wahl, 2008; Welsh, Ordóñez, Snyder, & Christian, 2015). All this underscores the need to find effective ways to regulate and reduce small-scale unethical behavior.

Fare evasion (or fare dodging) is one example of unethicality and is defined as the unlawful use of transit facilities by riding without paying the applicable fare (Bureau of Transportation Statistics).\textsuperscript{2} Fare evasion consists of travelling on public transportation without purchasing a ticket or without validating one’s ticket in the machine. According to the few studies that have recently examined fare evasion, it constitutes a major source of loss of revenue for transport operators (Clarke, Contre, & Petrossian, 2010; Currie & Delbosc, 2017; Delbosc & Currie, 2016; Guarda, Galilea, Handy, Muñoz, & Ortúzar, 2016). Weerman (2007) reported that half of all students in the Western urbanized part of the Netherlands admitted to fare dodging at least once during the preceding school year. In France, fare evasion rates are estimated depending on the type of transportation – to be between 5% and 25%, and costs operators about €600 million annually. This equals the cost of 2390 busses or the operating expenses for the public transit facilities of Lyon, Marseille, Toulouse, Lille, and Bordeaux combined. In addition to its direct costs, the costs associated with combatting fare evasion in France (such as hiring ticket inspectors or investing in new technologies such as ticket gates) represents roughly €195 million annually. In addition, findings indicate that fare evasion rates are related to criminal

\textsuperscript{*} Corresponding author.
\textsuperscript{1} E-mail addresses: s.ayal@idc.ac.il (S. Ayal), jeremy.celse@bsb-education.com (J. Celse), ghochman@idc.ac.il (G. Hochman).

https://doi.org/10.1016/j.obhdp.2019.10.003
Received 31 May 2018; Received in revised form 1 October 2019; Accepted 15 October 2019
0749-5978/ © 2019 Elsevier Inc. All rights reserved.

Please cite this article as: Shahar Ayal, Jérémy Celse and Guy Hochman, Organizational Behavior and Human Decision Processes, https://doi.org/10.1016/j.obhdp.2019.10.003
offenses such as vandalism or assaults against fellow passengers or staff, and measures to combat fare evasion have led to a significant decrease in these criminal activities (Smith & Clarke, 2000). Studies have also shown that increased rates of fare evasion generate perceptions of mistrust and insecurity, thus increasing passengers’ psychological discomfort and discouraging them from using public transportation (Bijleveld, 2007; Killias, Scheidegger, & Nordenson, 2009; Reddy, Kuhl, & Lu, 2011).

Traditional forms of regulation such as the standard economic approach to curbing fare evasion rely on external enforcement and deterrence. This approach tends to view passengers (fare evaders included) as rational agents who aim at maximizing their utility by exploiting the benefits of travelling for free against the costs of being caught (Becker, 1968; Boyd, Martini, Rickard, & Russell, 1989; Hirsch, Bottoms, Burney, & Wikström, 1999). Based on this assumption, fare evasion can be reduced by (1) increasing the likelihood of being caught (e.g., more ticket inspections, physical access barriers, video surveillance); and (2) increasing the severity of the punishment by imposing heavier sanctions. In fact, transportation companies in France, for example, have increased ticket inspection rates as well as the number of fines (e.g., in 2016, 97% of all transportation companies increased their fines), and even curtailed the presumed advantages of fare evasion (e.g., shorter period of ticket validity). Studies that have investigated external enforcement interventions stress the effectiveness of increasing the probability of being caught, and indicate that higher inspection rates tend to discourage fare evasion (e.g., Delbosc & Currie, 2016; Kooreman, 1993).

Nevertheless, recent studies have highlighted the limitations of external enforcement (e.g., Ay, Gino, Barkan, & Ariely, 2015; Hochman & Erev, 2013; Thaler & Ganser, 2015). More specifically, inspections are highly costly and it is virtually impossible to monitor the actions of all passengers at all times. Furthermore, external punishment and rewards can encourage on internal motivation (Ryan & Deci, 2000) and unethereal people from their own moral compass (Ariely, 2006; Bauman & Gino, 2012). Research on external inspections has reported mixed outcomes. Killias et al. (2009) for instance examined the effects of increasing the probability of inspections in Zurich's suburban transport system and found that fare evasion was negatively correlated with inspection rates but in a non-linear fashion. Similarly, Guarda et al. (2016) showed there was an optimal inspection rate above which inspection costs exceed the benefits. Other studies have reported no correlation between ticket inspection rates and fare evasion (Clarke et al., 2010). In terms of the severity of punishment; i.e., the sanction incurred if caught, Guarda et al. (2016) found that inspection strategies can be cost-effective even when evaders are not given a fine, and Bijleveld (2007) showed surprisingly that increased severity of punishment led to higher rates of fare evasion.

Here, we put forward two techniques of internal enforcement as an alternative approach to curb fare evasion rates, and describe how they were tested in the field. The basic theoretical premise behind our approach is that people value honesty and morality (Chainsen, Giner-Sorolla, & Chen, 1996; Greenwald, 1980; Sanitioso, Kunda, & Fong, 1990), and strive to maintain a positive self-image as moral actors (Duval & Wicklund, 1972; Rosenberg, 1979). At the same time, when there is an opportunity to benefit from immoral behavior, most people choose to do so, at least to a certain extent (Gino, Ayal, & Ariely, 2009; Hochman, Glückner, Fiedler, & Ayal, 2016; Mazar, Amir, & Ariely, 2008; Ordóñez & Walsh, 2015; Ordóñez, Schweitzer, Galinsky, & Bauman, 2009; Shalvi, Handgraaf, & De Dreu, 2011). This constant conflict between people’s unethical behavior and their desire to maintain a moral and positive self-image creates a form of psychological tension dubbed Ethical Dissonance (Ayal & Gino, 2011; Barkan et al., 2015). This tension may act as a moral gate-keeper that enables people to uphold their moral standards (Barkan et al., 2015). However, since people find creative ways to dismiss this dissonance, in many situations which pose a conflict between the desire to increase personal profit by cheating and the desire to maintain a positive self, cheating behavior prevails.

A growing body of research illustrates this dynamic, and shows that people take advantage of specific contextual factors and employ self-serving biases and justifications that enable them to benefit from dishonesty without negatively updating their moral ledger (Ayal, Hochman, & Ariely, 2016; Gino & Pierce, 2010; Shalvi, Gino, Barkan, & Ayal, 2015). For instance, Gino, Ay, and Ariely (2013) found that when people’s dishonesty benefits others, they are more likely to view dishonesty as morally acceptable, and thus cheat more and feel less guilty about it. In a similar vein, cheating level increases when people are confronted with more ambiguous situations (Hochman et al., 2016; Shalvi, Dana, Handgraaf, & De Dreu, 2011), or when they are exposed to wealth cues such as stacks of dollars (Gino & Pierce, 2009). Thus, a promising avenue of internal enforcement stems from the fact that the same factors that facilitate wrongdoing by reducing ethical dissonance can be harnessed to design policies and interventions that stress the dissonance and increase moral behavior. Specifically as regards socially desired behaviors, previous research points at two main factors that may help to design social environments that encourage compliance and moral behavior; informing people of the desired social norms and what others are doing (e.g., Bicchieri, 2005; Cialdini, Kallgren, & Reno, 1991), and giving people the feeling that they are observed by others (e.g., Ernest-Jones, Nettle, & Bateson, 2011; Nettle, Nott, & Bateson, 2012).

The first factor, which is the main focus of our investigation, is social proof (Cialdini et al., 1991; Reno, Cialdini, & Kallgren, 1993), a simple principle which argues that knowing what others are doing (descriptive norms) or what others think we should be doing (injunctive norms) exert enormous influence on behavior (Bicchieri & Dimart, in press; Cialdini & Trost, 1998). In one classic natural field experiment, Cialdini et al. (2006) documented the impact of descriptive social norm messages about the theft of petrified wood in a US National Park and indicated that a positive framing (e.g., percentage of people who do not steal) was more effective than negative framing (e.g., percentage of people who steal) in reducing theft. In a like manner, people are more likely to use the stairs rather than the elevator if a sign informs them that most people use the stairs than if the sign informs them about the health benefits of using the stairs (Burger & Shelton, 2011). Houten and Nau (1981) showed that informing drivers about the percentage of drivers who kept within the speed limit was ten times more effective than extra police patrols in reducing unsafe driving and speeding on the highway.

More recently, social norms have been used to successfully influence taxpayers to pay their taxes on time (Hallsowrth, List, Metcalfe, & Vlaev, 2017, see also Wenzel, 2005), citizens to reduce energy consumption (Allcott, 2011) and increase recycling (Cialdini, 2003). In addition, social proof was found to increase charitable giving (Frey & Meier, 2004) and voting participation (Gerber & Rogers, 2009), and improve financial decisions (Duflo & Saez, 2003; Karlan, 2007). Taken together, these studies suggest that the use of positive descriptive norms is more effective than traditional incentive approaches (that either reward desired behavior or sanction undesired behavior) and remain effective over time (Halpern, 2015; Sandburg, Schoenecker, Sebastian, & Sofer, 2009).

The second factor that may increase internal enforcement of moral behavior is visibility; that is, the feeling of being observed by others (Ayal et al., 2015). Classic research in social psychology demonstrates that anonymity releases people from their moral shackles (Milgram, 1973; Zimbardo, 1973). Thus, using visibility cues aimed at reducing the sense of anonymity should elicit desirable norms. For instance, placing people in a room with a mirror (Batson, Thompson, Seuerling, Whitney, & Strongman, 1999; Beaman, Klenz, Diener, & Svannum, 1979; Diener & Wallbom, 1976) or a well-lit room (Zhong, Bohns, & Gino, 2010) increased self-awareness and reduced unethical behavior. In fact, people are so sensitive to the sense of being observed that even
seeing a picture of watching eyes can alter their behavior for the better.

There are numerous examples of how the image of watching eyes can promote pro-social behaviors and reduce anti-social behavior (for a review see Baillon, Selim, & van Dolder, 2013; Dear, Dutton, & Fox, 2019). The best-known experiment used a picture of eyes next to an honesty box used to pay for drinks in a coffee room in the university (Bateson, Nettle, & Roberts, 2006). When the eyes were displayed, employees’ payments tripled. Similarly, Ernest-Jones et al. (2011) showed that people were more likely to clear their own trays in the cafeteria if the messages asking them to do so included an image of watching eyes, and Bateson, Robinson, Abayomi-Cole, Greenless, O’Connor, and Nettle (2015) showed that people are less likely to litter by dropping leaflets on a university campus if the leaflets included eye cues. Finally, Nettle et al. (2012) showed that placing a poster including a verbal message “Cycle thieves we are watching you” together with watching eyes decreased bicycle theft on campus.

Nevertheless, investigations of the effects of eye cues on dishonesty in the field are not as straightforward as one can think, and a series of replication attempts and new experiments have failed to report consistent evidence for an eye cue effect (for a detailed review and meta-analysis see Dear et al., 2019; Northover, Pedersen, Cohen, & Andrews, 2017). For example, Cai, Huang, Wu, and Kou (2015) showed that while an image of watching eyes promoted generosity in the dictator game, it had no effect on morality. Similarly, Baillon et al. (2013) showed that the effect of eyes works only in situations in which the decisions also influence the outcomes of others. Inconsistent results emerged from two recent meta-analyses: Northover et al. (2017) found no evidence that eye cues affect generosity, but Dear et al. (2019) found that these cues reduced anti-social behavior. Taken together, these studies suggest that watching eyes might be effective for certain behaviors (e.g., anti-social behavior) and in certain environments (non-busy ones). Thus, while watching eye cues may be effective in reducing fare evasion in general, they might not be effective in a busy train station. Most importantly, their effectiveness might increase when watching eyes are associated with other social cues that emphasize the desired behavior or convey positive information about the desired social norm (Nettle et al., 2012; Schild, Heck, Scigala, & Zettler, in press).

1.1. The current research and working hypothesis

Our primary objective was to test whether exposing people to positive descriptive social norms together with the visibility cue of watching eyes can actually reduce the rate of fare evasion among passengers in real life settings. To this end, we conducted two field experiments in railway stations in France. Study 1 was a natural field experiment conducted in collaboration with official ticket control inspections to examine actual fare evasion, whereas Study 2 was a lab-like experiment conducted in the field to examine the dishonest behavior of passengers who were exposed to our manipulation.

To the best of our knowledge, this is the first attempt to test the impact of social proof and eye cues on fare evasion in the field. In addition, similar to recent studies (Cohn, Maréchal, Tannenbaum, & Zünd, 2019; Dai, Galeotti, & Villeval, 2018; Hanna & Wang, 2017; Kröll & Rustagi, 2017; Potters & Stoop, 2016) we combine observations from a natural field experiment and a more controlled experiment conducted in the field (i.e., an artefactual field experiment, Harrisson & List, 2004). Thus, the findings should shed light on the relationship between field data and laboratory measures of unethical behavior, and responds to the call to test behavioral nudge interventions in real life under randomized controlled settings (Hallsworth et al., 2017; Halpern, 2015; Pierce & Balasubramanian, 2015; Sunstein, 2014).

To further investigate the actual effects of descriptive social norms with watching eyes and watching eyes alone in reducing anti-social behavior (Dear et al., 2019), we used a poster with a picture of watching eyes as a control condition, whereas in the experimental condition we added information about positive descriptive social norm (i.e., the percentage of passengers who purchase and validate their tickets) to the watching eyes poster. Based on previous literature and a pilot study conducted in our lab⁵, we hypothesized that exposing passengers to a poster campaign depicting a picture of watching eyes along with a social norm message (i.e., the experimental condition) would reduce fare evasion and would be much more effective than exposing passengers to a poster campaign with a watching eyes picture alone (i.e., control condition).

2. Study 1

The aim of this study was to examine the effect of eye cues alone and eye cues with a social norm message about fare evasion behavior. We conducted a natural field experiment in train stations in France. Based on previous results (Hallsworth et al., 2017), we used descriptive rather than injunctive social norms. In addition, since associating normative behavior with a geographical location is assumed to increase the saliency of the normative message (Hallsworth et al., 2017; Torgler, 2007; Wenzel, 2004, 2005), we mentioned in the message that the behavior was associated with people using that particular train station.

Finally, the social norm message was phrased to correct any misperception (or self-justification) that fare evasion is high. Note that our social norm message was framed positively (i.e., the actual (high) rate of passengers who obey the law and validate their tickets) for both theoretical and ethical reasons. Previous studies have shown that when descriptive social norm messages are framed positively they are more effective in curbing fraudulent behavior than negatively framed messages (Galdini et al., 2006). In addition, given the arrangement with the train company we were extremely reluctant to frame any message negatively so as not offend passengers or officials.

2.1. Experimental design and procedures

The study took place in Occitanie (South Eastern France). In Occitanie there are 101 railway stations, with a mean number of 5675.29 passengers monthly (SD = 11399.78). After a detailed review of these stations, the railway company allowed us to have access to two.⁴ The selection of the railway stations was based on the following criteria: 1) both railway stations had to be in the same geographic area and a similar distance to the main regional train station (Toulouse Matabiau); 2) both railway stations had to be connected to the same central railway station (to make sure that passengers had similar reasons for train travel); 3) both railway stations had to be similar in terms of traffic (to have a similar number of independent observations in each station to enable comparisons between the stations); 4) in both railway stations the connection between the station and the final destination (central railway station) had to be direct (fare dodgers might be informed about the presence of ticket control operations in stations and disembark early); 5) the local setting of the railway station had to be compatible with the display of to the message campaign; i.e., passengers using the station needed to be able to be exposed (consciously or unconsciously) to the messages; and 6) no other modes of transportation except automobiles connect the location of the train stations to the

---

⁴In a laboratory experiment, participants took part in a task which allowed them to cheat. Half of the participants (n = 50) were presented with instructions with eye cues on the top, and the remaining half (n = 50) with no image on the top. We observed a certain extent of cheating but no significant differences between the two conditions.

⁵Initially we planned to conduct the study in one larger station at different periods of time, but because of the threats of terrorist attacks, the company decided against this and only authorized two smaller sites.

⁶Based on information provided by the railway company, there is an association between travel distance and fare evasion: most fare dodgers travel short distances. Thus, we needed train stations that were equidistant from a main station (Toulouse Matabiau).
central railway station (to avoid interference with other messaging campaigns used for example by bus companies).

Two railway stations met these requirements: Colomiers and Muret. Colomiers is the second largest municipality in the department and Muret is the fourth largest. The population of Colombiers is 38,695 and 25,207 in Muret (INSEE, 2016). Both railway stations are in the same geographical area (less than 20 km between the two stations). Both railway stations are connected to Toulouse Matabiau (the largest railway station and one of the two largest train stations in terms of traffic, have very similar passenger traffic. The five largest train stations represent 51.9% of the total passenger traffic in the region. There is only one way to get to the platforms in Colombiers and two in Muret. In both stations, there are infrastructures to post posters. Hence, the layout of the stations guaranteed that passengers would be exposed to the messages. The only difference between the two stations is that whereas the connection between the Colombiers and the main regional station (Toulouse Matabiau) is direct, there is one stop between Muret and Toulouse Matabiau. Nevertheless, the travel time to the central station (Toulouse Matabiau) is similar (20 min). Fig. 1 presents a map of the train connections between the two stations.

Two messages were used in the study. In the control eyes condition, the posters consisted of a black background with “watching eyes” cues but did not include social norm messages. In the experimental eyes with social proof condition, the poster was the same as in the control condition except that it included a social norm message. The following sentence was included below the eye cues: “In this station, 90% of all individuals purchase and validate their ticket”. Figs. 2A and 3A present the two messaging campaigns used in the study. Muret was randomly selected to present the control eyes condition whereas Colombiers was selected to present the experimental eyes with social norm condition. The fare evasion rate mentioned in the social norm message corresponded to the average annual rate of fare evasion during the last 3 years in Colombiers (9.83% that we rounded off to 90%) and converted to people complying with regulations.

In both stations, two posters were displayed at each entry to the station and one poster was displayed on each platform (there were two platforms in each station). Normally, advertising spaces are rented to private companies for advertising purposes for a given time period. It turns out that the advertising boards were available for a two-week period at the time of the study (June 2017 and February 2019, as outlined below). Thus, posters were displayed for the entire two weeks. Figs. 2B and 3B provide an illustration of the message campaign in the two stations.

Two weeks before placing the posters and two weeks after removing the posters, train officials conducted a ticket control upon boarding and alighting to obtain a measure of actual fare evasion in both railway stations and to examine whether the messaging campaign had an effect on fare evasion. During a four-hour period, train officials blocked the entrances and exits of the train when it reached the station and asked each passenger to present his or her ticket. Ticket inspectors only checked the Muret or Colombier - Toulouse Matabiau trains (in both directions). Passengers were defined as fare evaders if they tried to board or to exit the train without a valid ticket. Identified fare evaders were fined on the spot. The fine consists of a fixed amount for fraud and a variable penalty depending on the distance between the first stop and the station where the ticket control took place. At the end of the day, train inspectors computed the fare evasion rate based on the number of passengers controlled and the number of passengers identified as fare evaders. Both ticket controls operations took place in the afternoon from 14:00 to 18:00. Note that in general, ticket controls do not take place more than once a month and are operated by a single team of inspectors (around 5 people) in the whole region. For the needs of the study, the national directors of the train company agreed to double the number of ticket inspection operations for one month and for the two stations in the study.

Data on fare evasion were collected on two waves to corroborate the results of the study. In Wave 1, we displayed the posters for a two-week period during June and July 2017. In Wave 2, we displayed the posters during February 2019. In Both waves, ticket control operations were conducted two weeks prior to displaying the messages and two weeks after removing the messages. While due to organizational constraints that were beyond our control, in Wave 1 the inspection after the messaging campaign took place on different days for both the control and the experimental stations (Tuesday vs. Thursday), in Wave 2 all inspections took place on the same days in both stations before and after the campaign (Thursdays). Table 1 summarizes the timeline of the study (for a full description of the designs and data in the studies and pre-registration information see https://osf.io/exn9z & https://osf.io/7cenz in accordance).

2.2. Results and discussion

In Wave 1 (left column of Table 2), before exposure to the message campaign, the fare evasion in the Control Station was 9.97% and 6.27% in the Experimental Station. After the two-week messaging campaign, the fare evasion in the control station dropped to 9.24% and to 3.25% in the experimental station. Thus, fare evasion fell by 8.11% (a 1 percentage point loss) after exposure to eye cues and by 52.63% (a 3 percentage point) after exposure to eye cues with descriptive social norms. However, in the Control Station, the fare evasion rates were not significantly different before and after exposure to the poster (Fisher Exact Test, p = 0.416, one-tailed). Conversely, in the Experimental Station, the fare evasion rates were marginally different before and after the messaging campaign (Fisher Exact Test, p = 0.065, one-tailed).

However, the initial differences in fare evasion rates between both stations need to be considered. The difference in fare evasion rates before the messaging campaign approached significance for the Control Station vs. the Experimental station (Fisher Exact Test, p = 0.055, one-sided). However, the difference in fare evasion rates between stations was much more pronounced after the messaging campaign (Fisher Exact Test, p = 0.002, one-tailed). To summarize, the fare evasion rate before and after our manipulation was higher in the control station than in the experimental station, with a significantly stronger difference after the manipulation.
A similar pattern of results was observed in Wave 2. As can be seen in the second column of Table 2, before exposure to the message campaign, the fare evasion in the Control Station was 9.1%, and 9.8% in the Experimental station. After the two-week messaging campaign, the fare evasion in the control station remained at 9.1% but dropped to 7.04% in the Experimental station. Thus, fare evasion remained unchanged after exposure to eye cues but fell by 28.16% (a 2.7 percentage point drop) after exposure to eye cues with descriptive social norms. In the Control station, the fare evasion was not significantly different before and after the intervention (Fisher Exact Test, p = 0.530, one-tailed). Conversely, in the Experimental station the difference in fare evasion rates approached significance before and after the intervention (Fisher Exact Test, p = 0.056, one-tailed). Here again, before our intervention, the fare evasion rates between the two stations were not significantly different (Fisher Exact Test, p = 0.383, one-tailed). Conversely, after the intervention, the difference in fare evasion rates between the two stations approached significance (Fisher Exact Tests, p = 0.066, one-sided). The effect of social norms messages was clearer in that replication since the fare evasion rates in both stations were

### Table 1

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td>Ticket controls #1</td>
<td>Messaging Campaign #1</td>
<td>Ticket controls #2</td>
<td></td>
</tr>
<tr>
<td>Wave 2</td>
<td>Ticket controls #3</td>
<td>Messaging Campaign #2</td>
<td>Ticket controls #4</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 2. Messaging campaign used in the control condition of the natural field experiment (A), and an illustration of the message campaign in Muret (B).](image1)

Fig. 3. Messaging campaign used in the control condition of the natural field experiment (A), and an illustration of the message campaign in Colomiers (B).
initially (before our intervention) equal. Moreover, since in Wave 2 ticket inspections occurred on the same day in both conditions, the fact that we found the same pattern of results in both waves suggest that the different days in which the control inspection took place after the messaging campaign cannot explain the results.

The overall data from both waves thus reveal that before our intervention, the fare evasion rates between the control and experimental stations were not significantly different (Fisher Exact Test, \( p = 0.217 \), one-tailed) whereas they were significantly different after our intervention (Fisher Exact Test, \( p = 0.006 \), one-tailed). Thus overall, the eye cues alone did not have a significant impact on fare evasion whereas the messaging campaign combining eye cues and a social norm message did generate an impact. Supporting our hypothesis, eye cues with descriptive social norms reduced fare evasion, presumably by prompting passengers to realize that fare evasion is not acceptable behavior.

Importantly, data from the time before the first wave obtained from the ticket control operation lends further credence to our results. That is, after our intervention in Wave 1, fare evasion dropped from 6.27% to 3.25%. To ensure this decline in fare evasion was due to our intervention, we requested train company data from ticket control operations conducted before our intervention to determine whether this decline might represent a normal fluctuation in fare evasion rates. Table 3 presents the results from previous ticket control operations conducted in the Experimental Station. We obtained data from 11 ticket control operations conducted before our intervention in the Experimental Station. The average fare evasion rate, based on these 11 ticket control operations was 6.19% (SD = 1.15). We calculated the confidence intervals to determine whether the decline we observed in our study might be observed in the Experimental Station without any intervention. Based on a normal distribution, the 95% confidence intervals were [0.054; 0.069] and the 99.9% confidence intervals were [0.045; 0.078]. Thus, at a 99.9% confidence level, it would be very unlikely to observe a fare evasion rate below 4.5%. After our intervention, however, the fare evasion rate declined to 3.25%. This suggests that a drop in evasion without any intervention was highly unlikely.

3. Study 2

In Study 1, we conducted a natural field experiment in which we compared the fare evasion rate in two train stations: one in which passengers were exposed to eye cues and another station in which passengers were exposed to eye cues and a descriptive social norm message. We found that exposing passengers to eye cues with descriptive social norms, but not eye cues alone, reduced fare evasion rates. In Study 2 we further tested the impact of our messaging campaign by implementing a more controlled laboratory task that is traditionally used to measure dishonest behavior (Gerlach, Teodorescu, & Hertwig, 2019). In addition, based on previous research which demonstrated a spillover effect in moral behavior (Bicchieri, 2005; Gino, Norton, & Ariely, 2010; Pierce & Snyder, 2008), we wanted to determine whether a behavioral intervention aimed at reducing a specific immoral behavior by positively describing a desired social norm (i.e., percentage of passengers who do not validate their tickets) would only have a specific effect on reducing fare evasion rate, or a more general effect on dishonest behavior. For this purpose, we conducted an artefactual experiment in the experimental train station.

### 3.1. Experimental design and procedures

At the same time as Wave 1 in Study 1, we conducted an artefactual field experiment in the Experimental Station (ES). Whereas in the natural field experiment, train officials controlled tickets upon boarding and alighting, in the artefactual field experiment, passengers on platforms waiting for their train were asked to participate in a variant of the standard die-under-cup paradigm (Fischbacher & Föllmi-Heusi, 2013; Shalvi et al., 2011b). Participants were told that they could earn €0, €3 or €5 by rolling a six-face die. The design used by Dai et al. (2018) was implemented, since their variant only has three possible outcomes rather than six, which made it possible to collect information on dishonesty without losing too much statistical power (given the inability to estimate the number of actual participants in advance). The die is placed inside an opaque cup and there is a peep-hole lid on the top (see Appendix). The sides of the die are colored blue, green and yellow rather than showing numbers. Hence there are 3 possible outcomes (green, blue, and yellow) at equal probability (1/3). Participants could earn €0 by reporting green, €3 by reporting blue and €5 by reporting yellow. Since only the participants knew the outcome of the roll, the task provides a good opportunity to cheat in order to maximize personal gain. Statistically, the probability of reporting a specific color (green, blue, or yellow) is one in three; hence, 33.33 percent. Since people are expected to cheat only by a little bit (Ariely, 2012; Ayal et al., 2016; Gino et al., 2009; Shalvi et al., 2011b), we expect that cheaters who roll green (€0) will report blue (€3), and cheaters who roll blue will report yellow (€5). Therefore, green will be reported less than 33.33 percent of the times and yellow will be reported more than 33.33 percent of the times (while blue will be reported about 33.33 percent of the times). Such a linear trend can be considered an indication of lying (Fischbacher & Föllmi-Heusi, 2013; Shalvi et al., 2011b).

For safety reasons, the train company only allowed us to conduct the die-under-cup task in one station. We thus focused on the station
where the social norm messages were displayed; i.e., the study was conducted in Colomiers (Experimental Station). As in Study 1, we asked officials from the train company for permission to replicate the die task in Wave 2. However, because of safety threats at the time of Wave 2 (the “Gilets Jaunes” demonstrations converging on Toulouse and elsewhere) the train company refused to let unauthorized personnel be on the platforms. Thus, Study 2 was conducted only in Wave 1. Passengers were recruited a week after the first ticket control and a week before the second ticket control (before and after being exposed to the poster), in order to assess the impact of the messaging campaign on dishonesty. The two experimental sessions were conducted on two different days and at different time slots (Monday morning from 8:00 to 12:00 for the time before the exposure to the posters and Tuesday afternoon from 14:00 to 18:00 for the time after the exposure). Note that the weather was similar in both sessions (cloudy but not raining). In addition, several demographics were collected to make sure that the different samples did not differ as a function of the different days in which data were collected. Table 4 summarizes the timeline of the experiment.

Research assistants who were blind to the objectives of the study, asked passengers waiting for their trains whether they would participate in a short study. Assistants (2 males and 1 female) were recruited by the research team and were wearing safety jackets with the word “survey” (“enquête” in French) written on it so they would not be misconstrued as train inspectors. The experiment was presented as a study conducted by researchers from the university on decision making. The assistants explained the task, the decision rule and rolled the die three times in front of the participant to ensure the die was not damaged or loaded (if the die was damaged, assistants could replace the die immediately). Then, the assistants gave the participants the cup, a pen, the instructions and a reporting sheet, and moved away so participants could sit at a table at the end of the platform in private, roll the die, fill in the reporting sheet and answer the demographic questions. Participants were asked to roll the die only once and to report the outcome of the roll. They were paid in cash immediately after handing their sheet to the assistants (see appendix for instructions).

3.2. Results and discussion

We set \( N \) such that the probability of correctly rejecting the null hypothesis of no lying when the null hypothesis is false was at least 80%. Specifically, \( N \) was defined such that if we observed 50% of participants reporting yellow (color leading to the highest earning), the binomial test had a power of more than 80% to reject the null hypothesis of no lying. Using \( p = 0.3333 \) (predicted) and \( p = 0.5 \) (observed), the required sample size was 60. Under the current setting (\( N_{\text{green}} = 103 \) and \( N_{\text{yellow}} = 60 \), the power to reject the null hypothesis of no lying was 93.77% before the intervention and 76.87% after.

Table 5 presents the demographics from the die experiments conducted on platforms. The samples were balanced with respect to gender (no significant differences between the two samples, \( \chi^2(1) = 0.290, p = 0.590 \), the reported ages were relatively similar in both sessions (no significant differences between the two samples in a two-tailed \( t \)-test for independent samples, \( t(1.61) = 0.782, p = 0.435 \)), and the job profiles were similar in both samples (no significant differences when comparing the distribution of professional status between the two samples, \( \chi^2(3) = 1.312, p = 0.726 \)). Most respondents were students, a few were unemployed and the proportion of working adults was similar in both sessions. Finally, when asked about the probability of having their ticket controlled (either on board or on the platform), the participants in both sessions made similar estimates. Specifically, the participants expected inspections to occur relatively frequently (44.56% before and 40.33% after, with no significant differences between the two samples on a two-tailed \( t \)-test for independent samples, \( t(1.61) = 1.179, p = 0.24 \). Bucchiol, Landini, and Piovesan (2013) also found high and biased beliefs concerning the probability of being controlled among Italian bus passengers. In fact, ticket checks are much less frequent than expected (less than 5%).

Table 6 presents the results from the two experimental sessions (before and after displaying eyes cue with social norm messages) on the platforms. Before the messaging campaign, a chi-square test of goodness-of-fit was performed to determine whether the three colors were equally reported. The three colors were not equally distributed in the population, \( \chi^2(2, N = 103) = 12.660, p < 0.01 \). Theoretically, since the die was not loaded there should have been an equal number of €0, €3 and €5. In other words, 33.33% of the participants should have reported €0, €3 or €5. In line with our expected linear trend for cheating, we found that low payoffs (€0) were significantly under-reported and larger payoffs (€5) were significantly over-reported. Specifically, 22.33% of the participants reported €0 (two-tailed binomial test that the observed percentage differed from 33.33%, \( p < 0.05 \)), 28.15% of the participants reported €3 (two-tailed binomial test, \( p = 0.216 \)) and 49.52% of the participants reported €5 (two-tailed binomial test 33.33%, \( p < 0.01 \)). Thus, consistent with previous studies (Dai et al., 2018; Shalvi et al., 2011b), the participants behaved dishonestly by inflating the die readings, but not to the full extent. After the messaging campaign, however, a chi-square test of goodness-of-fit revealed that the three colors were equally reported, \( \chi^2(2, N = 60) = 0.900, p = 0.637 \). In other words, the distribution of reported colors did not differ significantly from a uniform distribution after the passengers were exposed to social norm messages. None of the reported die readings differed significantly from the theoretical value of 33.33% (two-tailed binomial test, \( p > 0.1 \)). To summarize, in contrast to the first experimental session, we found no evidence of dishonest behavior on the die game after passengers were exposed to posters of watching eyes and social norms.

Additional analyses further supported our hypothesis. When comparing the distribution of die readings before the messaging campaign to the distribution of die readings after the messaging campaign, there were significant differences, using a linear-by-linear association model (\( 2, N = 163 \) = 5.685, \( p < 0.02 \). In line with these results, comparison of the average payoff in both sessions showed that before the eye cues with a social norm campaign, participants reported on average €3.320 (SD = 1.981) (honest = €2.666), while after the campaign, the reported average was €2.516 (SD = 2.151). That is, after exposure to eye cues with a descriptive social norm message about fare evasion, participants reported a significantly lower amount than participants before exposure to the message (two-tailed \( t \)-test for independent samples, \( t(1.61) = 2.42, p = 0.17 \)). Thus, the results provide additional evidence for the impact of eye cues with a social norm message on reducing dishonesty, and suggest that it not only affects the behavior highlighted

---

Table 4

<table>
<thead>
<tr>
<th>Week</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Die game before</td>
<td>Messaging Campaign</td>
<td>Die game after</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5

Demographics from the experimental sessions.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of females</td>
<td>45.63%</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>30.718 (15.592)</td>
</tr>
<tr>
<td>Proportion of unemployed</td>
<td>4.85%</td>
</tr>
<tr>
<td>Proportion of students</td>
<td>46.60%</td>
</tr>
<tr>
<td>Proportion of working adults</td>
<td>30.10%</td>
</tr>
<tr>
<td>Perceived probability of ticket check (SD)</td>
<td>44.563 (22.438)</td>
</tr>
</tbody>
</table>
Table 6
Descriptive statistics and results from two-tailed binomial test (that differed from 33.33%) results in the Die Experiments.

<table>
<thead>
<tr>
<th>Payoffs</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23 (22.33%)**</td>
<td>23 (38.34%)</td>
</tr>
<tr>
<td>3</td>
<td>29 (28.15%)</td>
<td>17 (28.33%)</td>
</tr>
<tr>
<td>5</td>
<td>51 (49.52%)***</td>
<td>20 (33.33%)</td>
</tr>
<tr>
<td>Nh Obs</td>
<td>103 (100.0%)</td>
<td>60 (100.0%)</td>
</tr>
<tr>
<td>Mean Payoff (Std. Deviation)</td>
<td>3.320 (1.981)</td>
<td>2.516 (2.151)</td>
</tr>
</tbody>
</table>

Significance levels from binomial tests (that differ from 33.33%) are coded as ** p < 0.05 and *** p < 0.01.

in the message, but moral behavior in general.

4. General discussion

The rapidly growing field of behavioral ethics is frequently being criticized for its overuse of abstract tasks in laboratory settings with low ecological validity to measure cheating behavior (where people know their behavior is being observed) as well as its focus on small scale cheating with modest financial stakes (e.g., (Bazerman & Tenbrunsel, 2012; Cohn, Maréchal, Tannenbaum, & Zünd, 2019; Henrich, Heine, & Norenzayan, 2010)). To address this criticism, we examined whether the insights gleaned from behavioral ethics research could affect moral behavior in the field, and whether the results obtained in the lab for the dynamic of small-scale cheating behavior could be replicated in real life. Two studies were conducted in two railway stations within the same geographic area in France. In Study 1, a natural field experiment, passengers were exposed for a two-week period to either a control eye-cue poster and an experimental eye-cue with a social norm messaging campaign. In Study 2, an artefactual experiment in the field, participants in the experimental train station were asked to participate in a variant of the die-under-cup paradigm (Dai et al., 2018) before and after they were exposed to the messaging campaign. Results from two waves of data collection in the first study with standard inspection operations showed a decrease in fare evasion in the experimental train station, whereas there were no changes in fare evasion rates in the control station. These results suggest that although the watching eye cues alone are not effective in a crowded train station, exposing passengers to watching eye cues together with a descriptive social norm messaging campaign could be an effective intervention to combat fare evasion.

Corroborating these results, the second study showed that people lied in the die task before exposure to the messaging campaign in the experimental station but not after. Thus, the findings across the two studies strengthen the external validity of laboratory experiments that document cheating behavior (Alem, Eggert, Kocher, & Ruhinduka, 2018; Dai et al., 2018, Potters & Stoop, 2016) since they demonstrate that outcomes derived from a widely used laboratory task to measure dishonesty (e.g., Abeler, Nosenzo, & Raymond, 2019; Bassarak et al., 2017; Cadshy, Du, & Song, 2016; Jacobsen & Piovesan, 2016; Mitikidis et al., 2017; Shalvi et al., 2011a, 2011b, 2012; Shalvi & Leiser, 2013) yield similar conclusions as field measurements obtained by standard ticket inspection operations. Moreover, the results imply that the level of dishonesty is not domain specific (Ariely, 2012; Ayal & Klar, 2014; Bazerman & Gino, 2012). Thus, our results support the growing body of literature on behavioral spillovers (Galizzi & Whitmarsh, 2019; Peysakhovich & Rand, 2015) and show that subtle manipulations that are aimed at curbing specific immoral behavior may yield a general immediate effect on moral behavior.

As stated in the introduction, findings are inconsistent for eye cues as a way to deter immoral behavior (e.g., Northover et al., 2017; Dear et al., 2019). Here as well, we found no effect for eye cues alone on fare evasion behavior. Thus, it might be argued that in a busy and crowded real life situations (such as a busy train station), eye cues are not sufficient to encourage ethical behavior. On the other hand, when the eye cues were presented together with a message drawing attention to the high rate of people who obey the rules, this messaging campaign was effective in reducing both fare evasion rates and dishonesty in general. This successful reduction of unethical behavior also supports previous findings on the effectiveness of social norms (e.g., Cicchieri & Dimant, in press; Gialdini et al., 2006; Hallsworth et al., 2017). As shown here, an effective social norm message should be descriptive (Hallsworth et al., 2017), self-engaging (Ayal et al., 2015), and positively framed (Gialdini et al., 2006). Interestingly, our results also show that some interventions that were found to be effective in laboratory settings (i.e., eye cues) might not work in specific environments in the real world. Thus, in order to increase the likelihood that behavioral interventions can affect moral behavior in real-life settings, different tools and different combinations of interventions need to be explored. Multiple nudges may be more efficient in direct people toward a desired ethical behavior.

Our study presents a natural field experimental investigation of eye-cues alone and eye-cues with social norms on fare evasion. One limitation of our study stems from the fact that when working with companies in the field, researchers cannot control all variables or implement optimal research designs. For example, since the railroad company only allowed us to run the experiment in two small stations, and to present our posters for two-week periods, we could only use two conditions. Moreover, we could not control the time or frequency of the inspection or the number of checked passengers. For that reason, we had to rely on fare evasion data that was collected 2 weeks after the posters were removed, and not during the messaging campaign, a limitation that might have underscored the power of the effect we found. However, since this constraint was implemented in both stations, it cannot account for the difference in effectiveness between the two posters. Still, additional research is needed to examine whether the effect obtained in our field experiment resulted from the combination of eye cues and descriptive norms, or whether descriptive norms alone would be sufficient to obtain this effect. Another limitation is the fact that the current study aimed at reducing behavior that has a relatively low baseline (less than 10%). On one hand, this worked to our advantage since we could use the actual percentage of people who follow the rules as the norm in the campaign message. On the other hand, it limited the power of our study and its significance level given the small percentages of fare evaders in our samples. Thus, we should restrict the generalizability of the eye cues and social norm intervention to less frequent phenomena, rather than more prevalent ones (e.g., tax evasion, intellectual property theft, etc.; see Cohn et al., 2019; Gino & Ariely, 2016), and strive to increase the statistical power of our examination in future research. Yet, the fact that we found an effect for eye-cues and social norm message with these limitations suggests that our manipulation might be even more effective in actuality.

Finally, since we were not able to debrief and question the train passengers after the manipulation, we can only speculate as to the potential mechanisms underlying the effectiveness of our messaging campaign in the experimental station. Still, the results from the questionnaire in Study 2 hint that the effect did not stem from external enforcement since most individuals overestimated the prevalence of inspections and the perceived probability of being caught was not significantly different before and after the messaging campaign. Thus, it cannot be claimed that fare evasion dropped because the passengers thought that there would be more enforcement following our messaging campaign. While external enforcement might explain the relatively low baseline of fare evasion, it might not be the most effective modus operandi in terms of cost-benefits, or its potential to further enhance passengers’ ethical conduct. Ticket control operations are costly, require the presence of several train officials who check potential transgressors at high frequency (Hochman & Erev, 2013), and can cause passengers to miss their connections (the train can only leave the platform when the ticket control inspection is over). Moreover, these operations generate stress that displace internal responsibility and may result in
aggressive behavior (Ariely, 2012; Ryan & Deci, 2000; Tyler, 2006).

To conclude, fare evasion is an example of small-scale dishonest behavior that causes considerable societal damage. In two studies, we demonstrated the potential of internal enforcement to combat fare evasion without increasing external enforcement or the severity of punishments. These interventions were inspired by the REVISE framework (Ayal et al., 2015) which defines three main principles underlying policy interventions to encourage moral behavior. In a recent study (Schild et al., in press), the authors tested the effectiveness of interventions that manipulate more than one principle, and specifically identified the effectiveness of a combination of Visibility cues and Self-Engagement as a way to achieve the maximum effect in reducing dishonesty. Presumably, eye cues combined with a descriptive social norm manipulation fit this category. Increasing the feeling that one is being observed and clearly defining the desired behavior as well as the (high) frequency of people who comply in this specific situation may stress the ethical dissonance associated with unethical behavior, enable people to connect the decision to their moral compass, and help them adhere to their moral values.

Acknowledgments

We are extremely grateful to Maud Bailly, Tanguy Cotte-Martinson, Joel Maillos, Antoine Marie and Jacques Breyssse for authorizing the project and contributing to its development. We thank Daniel Nettle and Melissa Batesson for sharing their research material. Comments from Mélanie Heugues, Wolfgang Steinel and Eli Spiegelman were particularly useful. Last but not least, we would like to thank Jean Einhorn, Isabelle Quinaux and David Paillisson for their help in administering the experiment and collecting the data. This study was funded by the SNCF.

Appendix

Pictures of the die-under-cup used in the artefactual experiment.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.obhdp.2019.10.003.

References


Alem, Y., Eggert, H., Kocher, M. G., & Ruhinduka, R. D. (2018). Why (Schild et al., in press), the authors tested the effectiveness of interventions to encourage moral behavior. In a recent study work (Ayal et al., 2015) which demonstrated the potential of internal enforcement to combat fare evasion without increasing external enforcement or the severity of punishments. These interventions were inspired by the REVISE framework (Ayal et al., 2015) which defines three main principles underlying policy interventions to encourage moral behavior. In a recent study (Schild et al., in press), the authors tested the effectiveness of interventions that manipulate more than one principle, and specifically identified the effectiveness of a combination of Visibility cues and Self-Engagement as a way to achieve the maximum effect in reducing dishonesty. Presumably, eye cues combined with a descriptive social norm manipulation fit this category. Increasing the feeling that one is being observed and clearly defining the desired behavior as well as the (high) frequency of people who comply in this specific situation may stress the ethical dissonance associated with unethical behavior, enable people to connect the decision to their moral compass, and help them adhere to their moral values.


