# HONESTY ON THE STREETS: A FIELD STUDY ON NEWSPAPER PURCHASING

**Gerald J. Pruckner**
University of Linz

**Rupert Sausgruber**
Vienna University of Economics
and Business

**Abstract**
Many publishers use an honor system for selling newspapers in the street. We conducted a field experiment to study honesty in this market, finding that a moral reminder increases the level of honesty in payments, whereas the same message has no effect on whether one is honest. Reminding customers of the legal norm has no effect. We argue that these results are consistent with a preference for honesty, based on an internalized social norm. Auxiliary evidence suggests that the moral message remains effective when it is posted for longer periods, and even when it is removed again. (JEL: C93, K42)

## 1. Introduction

Researchers in economics have been giving increased attention to the importance of norms—in particular, to honesty (see, among many others, Bénabou and Tirole 2006). Honesty norms are likely to play a key role in the occurrence of corruption (Rose-Ackerman 2001; Barr and Serra 2010) and of white-collar crime in the guise of employee theft (Greenberg 1990), tax compliance (Erard and Feinstein 1994), insurance fraud (Picard 1996), or managerial fraud (Mittendorf 2006).[1] However, it is difficult to isolate the effect of such norms from alternative accounts.

In this paper, we present the results of a field experiment and two complementary studies that together address the role of honesty norms among customers in a real

1. Mazar and Ariely (2006) provide an excellent overview of the importance of internalized reward mechanisms for the occurrence of honest and also dishonest behavior.

FIGURE 1. Newspaper sales booths with plastic bag for the papers and padlocked cashbox. Helpers mounting a sales booth onto a light pole (left) and a booth showing the Moral treatment's message (right).

market for newspapers where payments are not monitored. In order to generate an exogenous variation in norms, we post a message for customers that reminds them to be honest; we then evaluate the effect of this message against a neutral control. We also compare the effects of the moral reminder to a reminder of legal enforcement. This approach, which is based on work in social psychology (see, for example, Schultz et al. 2007), allows us to obtain qualitative evidence on the role of norms by comparing behavior among members of the moral reminder group and behavior of those in the control group. Comparing the effects of different message types enables us to evaluate the relative strength of an explanation that is based on internalized social norms against one based on compliance due to external forces (see Cialdini and Goldstein 2004).

Austrian print production companies commonly sell tabloids on the streets, via an "honor system" that involves a booth containing a bag filled with newspapers and a padlocked cashbox (see Figure 1). The customers are supposed to deposit payment into the cashbox, but they may also underpay or simply take the paper without paying.[2] We use this setup to record the behavior of people who encounter an opportunity for dishonest gain. The gain is petty, but the probability of being caught is negligible and

2. For a splendid description of an honor system, see Dawes and Thaler (1988). Their example of how local farmers once sold produce next to the road has become a renowned case in behavioral economics. Our setting resembles that "pay what you want" system, but with the important difference that, in such a system, those who fail to make a voluntary payment are not regarded as being dishonest (see Gneezy et al. 2010).

the opportunity is tempting. Hence, the setup provides a unique opportunity to study determinants of honesty under controlled conditions in the field.

Our experiment involves imprinting messages for the customers on the sales booths, using two treatments and a control that were randomly allocated to booths. In the so-called Legal treatment, we remind customers of the legal norm by posting the message that stealing a paper is illegal. The message in the Moral treatment aims to reinforce customer honesty by activating a social norm (and behavior) that benefits society. The experiment was conducted on three consecutive days in two towns in western Austria. In total, we observed 120 instances of customers taking a paper; in each case, we checked to see how much was paid for it.

The results reveal that two-thirds of customers do not pay for the newspaper, and those who do pay deposit much less than the indicated price of €0.60. On the extensive margin, the treatments have no effect on whether one is honest. The striking result is that customers do react on the intensive margin to the appeal for honesty: the average payment per newspaper in the Legal and Control treatments is less than a tenth of the actual price, whereas the average payment in the Moral treatment is nearly a fourth of the price. This positive effect (for those who pay some positive amount) of appealing to honesty is consistent with a preference for honesty that is based on an internalized social norm. When cheating is materially beneficial, those who live up to this norm face a conflict with self-interest; the literature proposes that people in such situations can reduce cognitive dissonance by engaging in self-deception (Festinger 1957), thereby creating moral "wiggle room" that enables them to act in a self-serving manner. In a study related to ours, Mazar, Amir, and Ariely (2008) perform laboratory experiments in which a moral reminder reduces the extent of such self-deception. Our study supports those findings with field data.

We conduct two auxiliary studies to complement the field experiment. The first one aims to shed light on the questions of how the same customers react to an appeal for honesty when it is posted for a longer period and how they react when the appeal is removed. We collect daily data on revenue and the number of papers taken per booth from the same 250 locations over seven consecutive weeks. After three weeks, we post a moral reminder in a subset of 127 locations; after two more weeks, we remove the reminder. In line with our results from the field experiment, we see an increase in revenue shortly after posting the Moral message. In contrast, revenues in locations without a reminder remain almost unchanged. The figures give no indication that the reminder effect would diminish over time, and it does not break down soon after the message is removed. Yet because the locations were not randomized in this part of the study, we must interpret these results with care.

The second auxiliary study combines the results from a survey with examinations of individual payments. The procedures are such that the customers are unaware that their payments have been verified. We can hardly claim causality, but this setup does allow us to identify sociodemographic correlates of honesty. We find that males pay less than females and that people living with partners pay significantly more than singles. Moreover, honesty is positively correlated with several variables: the individual's trust in the legal system, whether a respondent volunteers, a measure of tax compliance,

and the stock a person puts in others' opinions about him. However, we find a negative correlation between honesty and church attendance.

Our work relates to several strands of the extant literature. Haan and Kooreman (2002), Bateson, Nettle, and Roberts (2006), and Levitt (2006), all provide field evidence that is consistent with internal rewards for being honest. Moreover, a recent wave of laboratory experiments has established the existence of a preference for honesty (Fischbacher and Heusi 2013; Mazar, Amir, and Ariely 2008; Gibson, Tanner, and Wagner 2013; Houser, Vetter, and Winter 2012). The similar findings of Bolton, Katok, and Zwick (1998), Konow (2000), and Dana, Weber, and Kuang (2007) indicate that people attempt to camouflage self-serving choices in a way that makes them appear unselfish. Honest behavior has also been observed when people interact strategically (Brandts and Charness 2003; Gneezy 2005). Focusing on this aspect of strategic interaction, several authors have modeled honest behavior as a preference for fulfilling others' expectations regarding outcomes (e.g., Charness and Dufwenberg 2006). On the other hand, Vanberg (2008) reports results that suggest people have also internalized a norm for promise keeping per se. Such behavior is captured by models of identity and self-image in which people value themselves for being honest (e.g., Bénabou and Tirole 2006; Ellingsen and Johannesson 2008; Andreoni and Bernheim 2009).

Our study extends the existing literature along three dimensions. First, we employ field data generated under controlled conditions and within a truly natural context. In doing so, we overcome the main disadvantage of laboratory studies—namely, that subjects may change their behavior in response to being observed by the experimenter. Second, we focus on evaluating the effect of a message as a simple, nonprice intervention to persuade customers to be honest. This method provides causal evidence pertaining to the role of social norms. Finally, we use survey data to help identify individual predispositions to honesty based on observable signals and personal characteristics.

The rest of this paper is organized as follows. Section 2 describes the empirical setup and the methods used to generate the data for our studies. Section 3 reports the results, and Section 4 summarizes our findings and gives concluding remarks.

## 2. Setup, Data, and Procedures

We study the behavior of customers in a market that sells tabloids without monitoring payments. The sellers position hundreds of sales booths near the streets; each booth consists of a plastic board that is $25 \times 15$ inches in size, a moisture-proof plastic bag from which customers take the paper, and a cashbox. The price of the paper is indicated on the cashbox.

Although the sales-booth system is especially popular on weekends, many publishers also employ it on weekdays.[3] A question of immediate interest is *why*

---

3. Labor market regulation makes it expensive to hire labor on weekends, and shop hours are restricted. On Sundays, then, Austrians can shop only in a few places—for example, airports, railway and gas stations, and tourist centers. In the newspaper market, publishers often provide a weekend home delivery for those who subscribe to their "serious" papers.

publishers employ the sales-booth system.[4] According to our data and private information from the publishers, revenue from sales alone would not make the system worthwhile. Instead, the lion's share of publisher revenue comes from selling print space to advertising purchasers. Because the selling price of this space is increasing in the paper's circulation, sellers can make up for pilferage by distributing more copies (see Picard and Lacy 1999).

In conducting this study, we collaborated with two tabloid publishers. The first operates in Vorarlberg, a province in western Austria with 350,000 inhabitants; the second is located in Upper Austria, a northern province with 1,410,000 inhabitants. Both hold large market shares in the regional tabloid market. They distribute much of the print run via the sales-booth system, but they also distribute papers through shops or directly to the homes of readers who hold subscriptions. In Vorarlberg at the time of the study (in 2004), the Sunday print run of the publisher's tabloid was 33,000; on weekdays, it was 25,000. The estimated number of readers in that year was 64,000, or 23.7% of all potential readers over age 14 in the province. In Upper Austria, the publisher's average daily print run, between July 2008 and June 2009, was 110,000 papers, which were distributed to 355,000 weekday and 397,000 weekend readers.

We collected data from the sales-booth system in three different ways. Our main study consisted of a *field experiment* in which we randomly assigned treatments to different booth locations.[5] At every location, we collected information on individual payments from customers who had taken a copy of the newspapers from the booth. In addition to the field experiment, we conducted a *quasi–field experiment* to collect aggregate data from a large number of locations over time. These data are only suggestive, but they allow us to observe the same customers' reactions to an appeal that is posted for a longer period. Finally, we posed *survey* questions to customers who had taken a copy of the newspaper from a sales booth. The survey procedures were such that we could—anonymously and without their knowledge—match the customers' responses with their payments. The field experiment and the quasi–field experiment each used an appeal to honesty as the main treatment, which differed from the control treatment. The survey aimed to identify correlates of honesty and was run under constant conditions; that is, there was no experimental variation during the survey. Although the three data collection steps were independent of each other, they all contributed to our understanding of why people are honest in this market. Next we shall provide further details of the field experiment and the quasi–field experiment. For more information about the survey, see Section 3.3 and Online Appendix B.

## 2.1. Field Experiment Details

Table 1 presents the exact wording of the message in each of the treatments. The messages were printed in large and clearly legible letters on the cover of the bag

---

4.  For a discussion of the different modes of newspaper circulation, see Bradshaw (2003).

5.  This approach is a *descriptive* field experiment under the classification proposed by Card, DellaVigna, and Malmendier (2011).

TABLE 1. Treatments of the field experiment.

| Treatment | Message |
| --- | --- |
| Control | *The paper costs €0.60.* |
| Legal | *The paper costs €0.60. Stealing a paper is illegal.* |
| Moral | *The paper costs €0.60. Thank you for being honest.* |

holding the papers (see Figure 1). A customer had to lift this cover in order to remove a paper from the bag, so we are confident that all the customers saw and read the message. We took extreme care in ensuring that the presence of experimenters did not affect customer anonymity; for this reason, the experimenter placed just one paper in the bag of the sales booth and checked for payments at intervals of approximately 40–60 minutes. If the paper had been taken from the bag, the experimenter opened the padlock, emptied the cashbox, and recorded the amount of payment. Then the experimenter refilled the bag—again with only one paper—and moved on to the next location. This procedure minimized the possibility that customers noticed the experimenter recording payments or felt they were being observed when they took the paper.

We ran the experiment during the week, initially on three consecutive days of June 2004 in a town with 44,000 inhabitants; the second time was on three consecutive days of October 2004 in a town with 28,000 inhabitants. In each town, we chose a set of potential locations: 21 in the first town and 19 in the second. The locations were close to the town centers, which are lively shopping and recreation areas. These locations had several desirable characteristics in terms of our study's purpose. First, they were frequently the sites of the sales-booth system and so customers were accustomed to obtaining papers there. Second, we wanted to collect as many observations as possible and chose locations that were highly frequented. Finally, to save time in collecting data, the locations in each set were within a reasonable walking distance of each other.

The publisher provided us with ten sales booths. On each day of the experiment, we randomly assigned the treatments to the locations and mounted a sales booth at each of them. Randomization was achieved in the following manner. On each day of the experiment, we sorted the potential locations in a town by using computer-generated random numbers drawn from a uniform distribution. We mounted a booth at each of the first ten locations, and we followed a treatment list that was prepared for that day. On each day, this list specified four locations for one treatment and three locations for two treatments; the respective treatment implemented at the four locations alternated daily. For example, on day 1 of the experiment, three locations were each assigned to the Control and Legal treatments and four locations to the Moral treatment. On day 2, we had three locations each for the Moral and Legal treatments and four locations for the Control treatment.

The data were collected between 8 a.m. and 1 p.m. each day. During this five-hour period, the experimenter walked from one location to the next and, after collecting the data from all locations, started all over again. Thus, over six days and in two towns

we checked the sales booths at 40 locations a total of 333 times and observed whether a copy had been taken and paid for. Each treatment was implemented 20 times. We selected 21 locations once for the experiment and 18 locations twice. Only one location was selected three times for the experiment. Because of the randomization, treatments did not differ with respect to whether they were implemented in locations that had been used repeatedly.[6]

## 2.2. Quasi–Field Experiment Details

The quasi–field experiment was run in a town of 190,000 inhabitants in Upper Austria. The data were gathered via the publisher's standard routine of mounting the sales booths and collecting the revenue, which we summarize as follows. The publisher operates on 13 routes, which involve between 44 and 130 locations each (the median is 57). For our study, we selected two routes that had many locations within a concentrated area and similar sociodemographic characteristics. The first route combines 127 locations over eight districts; the second route has 123 locations in five districts.[7] A coin toss decided that the second route would be used for the control message and the first for the honesty appeal.

The study spanned seven weeks, starting with the third week of January 2010 lasting through the first week of March 2010. Each week, the study was implemented on the same three weekdays, Wednesday through Friday.[8] In period 1 (weeks 1–3), the sales booths at all 250 locations posted the control message *The paper costs €1*.[9] In the control locations, this message remained posted for the entire length of the study. In the treatment locations, during period 2 (weeks 4 and 5) we replaced the control message with the following message: *The paper costs €1. Thank you for being honest*. In period 3 (weeks 6 and 7), we changed this back to the control message. Comparing the difference in outcomes over periods 1 and 2 between the treatment and control locations replicates the test of the field experiment in an alternative setup. In addition, examining the difference-in-differences (DID) of outcomes over periods 1 and 3 enabled detection of any delayed effects after the posting was removed.

Drawing inference from this design rests on the identifying assumption that the differences in outcomes over time would have been the same in the treatment and

---

6.    The number of locations that were used only once were seven for Control, six for Legal, and eight for Moral; the number of locations that were used twice were (respectively) 12, 13, and 11. The one location that was used three times was assigned to all treatments. The distribution of these numbers does not differ among the treatments ($p = 0.836$, Kruskal–Wallis test).

7.    The city has 36 districts in total.

8.    Were it not for the study, the publisher would not have used the sales–booth system during the week. To lessen the publisher's burden, we agreed to run the experiment on three consecutive weekdays. The publisher normally limits its use of the sales-booth system to distributing a weekend edition throughout Upper Austria. The weekend edition has a higher price (€1.50 versus €1.00), is richer in content, and has more supplements and advertisements than the regular weekday edition; and on the demand side, the weekend customers probably differ from those in our sample. For these reasons, we decided to exclude Saturdays from our study.

9.    At the time of the experiment, the actual price of the paper was €1.00.
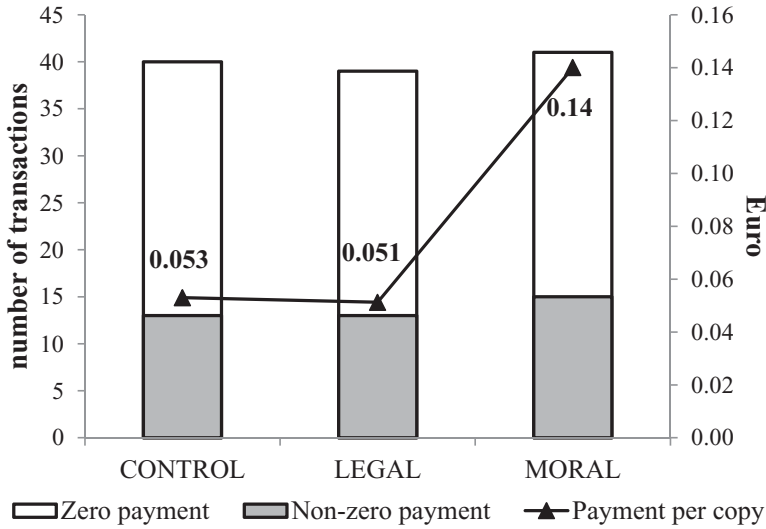
FIGURE 2. Number of transactions and payment per transaction (including zero payments) over treatments ($N = 120$).

control locations if the message had *not* been posted. On the one hand, our setup avoids some common pitfalls of uncontrolled natural experiments because the individuals being observed can neither self-select into the treatment nor change their behavior in anticipation of being treated. On the other hand, the locations were not randomized over the treatments and we cannot claim that the locations on any given route are independent. Hence, we interpret our findings with care.

## 3. Results

In Section 3.1, we present the outcomes of the field experiment. Sections 3.2 and 3.3 present (respectively) the results of the quasi-experimental study and the survey.

### 3.1. Field Experiment Results

Altogether, at 40 locations in six days we observed 120 instances in which the paper was removed from the bag. In the following account, we call the act of taking a copy being taken from the booth a "transaction". In 41 of 120 cases, a positive payment was recorded. Figure 2 breaks down the data by treatment type, and shows how many of the transactions resulted in no payment. The height of the white bars indicates the total number of transactions (e.g., 40 in Control); the gray bars indicate the number of transactions for which a positive payment was recorded (e.g., 13 in Control), so the difference between the white and gray bars is the number of transactions for which the payment was zero (e.g., 27 in Control). Finally, the height of the triangles with respect

to the right-hand vertical axis indicates the average payment (including nonpayers) per transaction (these values are also printed in boldface).

Our first observation based on Figure 2 is that almost two-thirds of customers do not pay for the paper. The low level of honesty in our field setting contrasts with reported laboratory evidence—for example, by Gneezy (2005), Fischbacher and Heusi (2013), and Mazar, Amir, and Ariely (2008). Because the experimental setups in these works differ from ours in numerous ways, there could be many reasons for this discrepancy.[10] Our second and no less important observation concerns the number of free riders: the frequency of dishonest customers is almost equal across the treatments. Figure 2 reveals that 67.5% (resp., 66.7% and 63.4%) of customers were free riders in the Control (resp., Legal and Moral) treatment; these numbers are not statistically different from each other.[11] Finally, the figure shows an increase in the average payment per copy in response to the appeal for honesty. The average payment is €0.053 in Control ($N = 40$), €0.051 in Legal ($N = 39$), and €0.14 in Moral ($N = 41$). An overall analysis of variance (ANOVA) reveals a significant effect of the Moral and Legal treatments ($F(2, 119) = 3.10$, $p = 0.049$). However, this effect is entirely due to the treatment Moral ($p = 0.035$),[12] as payments in treatment Legal do not differ from those in Control ($p = 0.967$).[13]

Figure 3 presents the frequency distribution of nonzero payments per treatment. The standard model predicts that customers will pay the full price of the paper (€0.60) only if the expected sanctions in the event of detection outweigh the benefit of stealing. The pattern graphed in Figure 3 runs counter to this prediction in that some people make positive payments that are still lower than the full price.[14] There is a pronounced spike in the distribution of small payments in the Control and Legal treatments, and the payment patterns under these two treatments are strongly similar. In contrast, payments in the Moral treatment shift upward; here, in fact, the distribution's mode is at the paper's full price. The average (nonzero) payment is €0.163 in Control and €0.154

---

10.    An obvious possibility is that people in a laboratory experiment react to being observed by others. In our study, we took great care to keep decisions fully anonymous. See Gerber, Green, and Larimer (2008) for a study showing how being observed can have dramatic effects.

11.    For example, a $\chi^2$ test indicates that the $p$-value between the Control and Moral treatments is 0.439.

12.    Our experimental procedure is such that multiple observations from the same customer are unlikely. In the first place, a potential buyer has no use for more than one paper per day. Second, since there is only one paper in a bag, and since each bag is restocked only at 40–60 minute intervals, it is unlikely that we will observe the same buyer more than once. Nevertheless, to account for this objection we performed another statistical analysis that excluded all observations from locations that were used more than once. Although this restriction reduces the number of observations from 120 to 37, the Moral treatment still remains significant ($p = 0.097$; the $p$-value for Legal is 0.538).

13.    If our randomization was successful, then the locations at which Moral was implemented should not differ from other locations. To test for this, we consider the payments at locations for which Moral was not implemented on a particular day ($N = 120 - 41 = 79$) and compare them across locations that had implemented Moral (i) never and (ii) at least once on another day. In case (i), $N = 54$ and the average payment is €0.059; in case (ii), $N = 25$ and the average payment is €0.038 ($p = 0.900$, Mann–Whitney test).

14.    We observe a few cases of customers who overpay: one payment of €0.70 in the Control treatment and three payments of €0.70 in the Moral treatment. A payment of €0.70 is most probably due to honest people who lack the appropriate coins to pay the exact price.
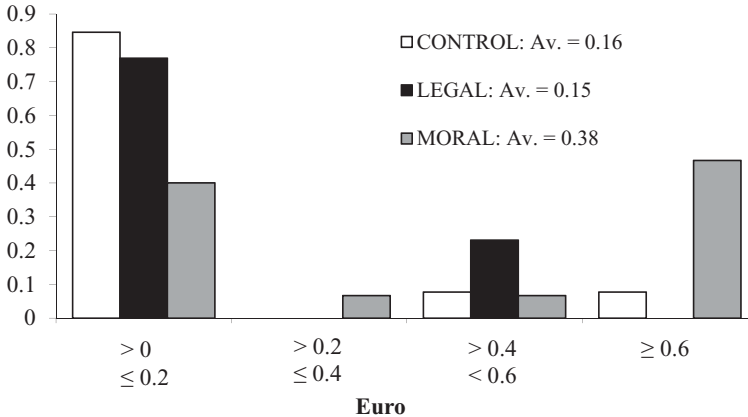
FIGURE 3. Distribution of nonzero payments by treatment ($N = 41$).

in Legal, whereas non-free riders pay €0.383 in Moral. The difference is significant between Moral and Control ($p = 0.033$, Wilcoxon rank sum test) and between Moral and Legal ($p = 0.008$). Comparing the nonzero payments jointly in all treatments, a Kruskal–Wallis test indicates that the data cannot be regarded as being sampled from the same population ($p = 0.017$).

It is important to note that our findings do not imply that legal deterrence has no effect. After all, if a legal reminder does not change behavior then this might simply indicate that people are already aware of the legal situation and the parameters of deterrence that apply.[15] In particular, the standard deterrence model does explain the behavior of most customers (i.e., those that did not pay) in that the absence of sanctions in our experimental context encouraged free riding. With regard to customers who make positive payments, the effects of appealing to honesty are consistent with a model in which people face a trade-off between the material benefits of stealing and adherence to an internalized norm of honesty.

Figure 4, which we have adapted from Cooter (1998), illustrates this interpretation. The figure orders the customers with respect to their willingness to pay so that they honor the honesty norm. If deterrence is absent and if choices are fully anonymous, then only those with a positive willingness to pay are said to have internalized the norm. The others pay zero because they respond only to external incentives, which are downplayed in our setup. The figure assumes that honesty is negatively associated with how much it costs to obey the honesty norm—in other words, that the fraction of buyers paying in full is decreasing in the price of the paper.[16] We observe that approximately one-third of the customers pay positive amounts, but only a few of

---

15. An alternative explanation is that the lack of enforcement in the past rendered our Legal treatment not credible to customers.

16. This accords with Levitt (2006), who finds a negative correlation between honest payments for bagels and their price.
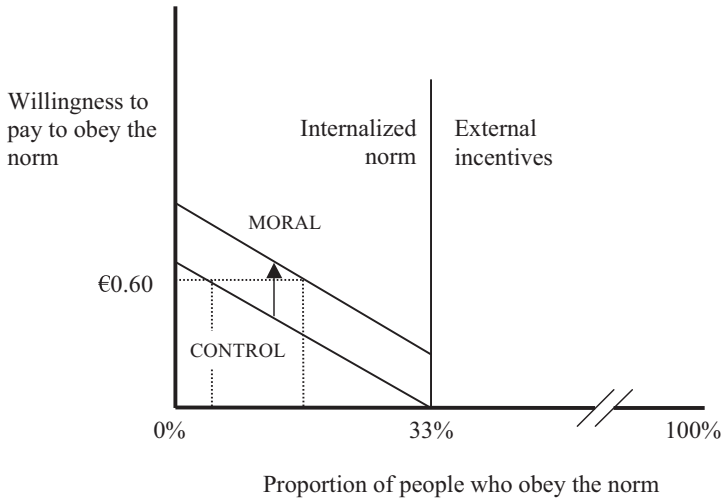
FIGURE 4.   Illustration of the treatment effect (based on Cooter 1998).

them place a sufficiently high intrinsic value on obeying the norm to pay the paper's full price. Moreover, two-thirds of the customers choose not to pay. We have seen that the message in the Moral treatment induces higher payments, yet it does not affect the frequency of zero payments. This pattern is consistent with an upward shift of the willingness-to-pay schedule to the left of the line (in Figure 4) that separates those customers who have internalized the norm from those who have not.[17]

Overall, our experiment provides evidence—from a natural field context—that is consistent with an internalized norm of honesty. Our observations are in line with the results of Mazar, Amir, and Ariely (2008) and Fischbacher and Heusi (2013), who report behavior in terms of a preference for honesty even in the absence of external incentives (such as those that prevail in laboratory experiments). In addition, Mazar, Amir, and Ariely find that the prevalence of honesty is not sensitive to the perceived likelihood of detection, although it does increase in response to moral appeals. Our own results are in accord with these findings.

### 3.2.  Quasi–Field Experiment Results

From 250 locations over seven weeks (three days per week), there were 12,985 transactions that generated total revenue of €627.74.[18] Given the price of €1 per copy, this money accounts for less than 5% of the revenue actually due (i.e., the

---

17.   Figure 4 serves mainly illustrative purposes; we assume (without explicit justification) that the willingness-to-pay schedule is linear in its positive range and shifts in a parallel fashion.

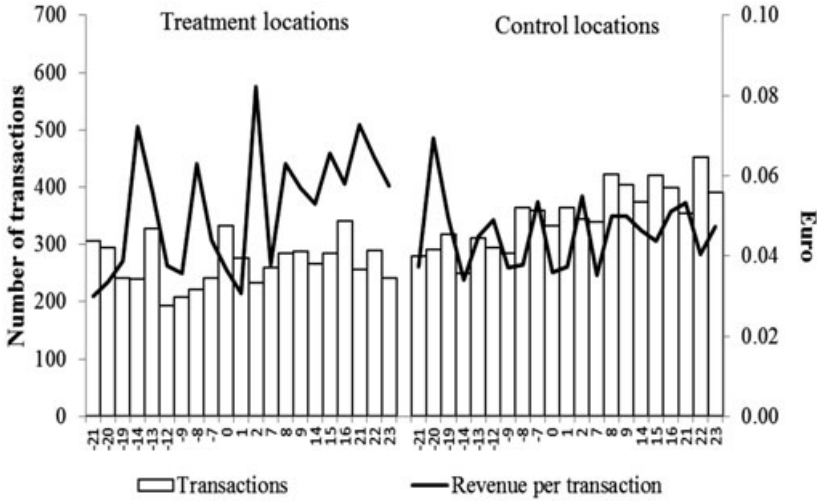18.   In 1,199 cases, no paper was taken and the cashbox was empty.

FIGURE 5. Daily number of transactions and daily revenue per transaction, by treatment locations ($N = 127$) and control locations ($N = 123$).

amount collected had all customers been honest in their payments). Thus, aggregate honesty was little more than half of what we observed in the field experiment.[19]

Figure 5 presents the time series of daily data by control and treatment locations. The time line shows calendar days, which are renormalized so that "0" is the first day of posting the appeal to honesty in the treatment locations. The height of the white bars on the left scale measures the daily number of transactions, and the bold plot (keyed to the right scale) indicates revenue divided by the number of transactions per day.

The evidence presented in this section is both suggestive and weak. It is suggestive because we cannot rule out the possibility that outcomes for sales booths along a route are correlated (see the discussion in Section 2.2). Hence there could well be reservations about the internal validity of the DID method. The evidence is weak because the data show substantial variation—notably, in the pre-experimental period. Furthermore, we observed in the field experiment that approximately two thirds of customers do not pay. Yet because aggregate honesty is much lower, the proportion of zero payments may be even greater in the quasi-experiment. Together with the main results of our field experiment, these low levels of honesty imply that it may be difficult to detect a treatment effect even with a large number of observations.

Although we interpret the data with care, the visual impression given by Figure 5 is in line with findings of the main field experiment. In the treatment locations, revenue

---

19. There, the average levels of honesty were 8.8% (€0.053/€0.60) and 8.5% (€0.051/€0.60) in the Control and Legal treatments, respectively (see Figure 2). We remark that, apart from other obvious differences, the two studies were conducted in different geographical regions; those regions may be characterized by different sociodemographics and different moral standards vis-à-vis honest behavior in such unmonitored payment situations.

"flares up" after the start of the treatment.[20] With the exception of a one-day trough (day 7), revenue remains high and there is no indication that the reminder effect will diminish over time. In contrast, revenue remains essentially unchanged in the control locations.

It is noteworthy that Figure 5 shows revenue remaining high in the treatment locations even after the message has been removed. First-time customers who enter the market during this time would not see the appeal; therefore, presuming the data do not drift apart in the absence of the treatment, the sustained higher revenue could only be due to customers who were previously exposed to the appeal for honesty and thus remain honest thereafter. It is not plausible that such effects would persist for a long time, but the pattern in Figure 5 suggests that there are no significant costs associated with introducing and then removing an appeal to honesty.

Let us next consider the number of transactions. Figure 5 shows an increase in transactions over time in the control locations (as indicated by the level of white bars in the right panel). From this observed trend in the control locations, we can reasonably infer that there is a (net) inflow of new customers into this market. Notice that revenues per transaction remain fairly constant over time, so incoming customers are not any more or less honest than existing ones. Although patterns differ among treatment locations, overall the number of such transactions varies little from one week to the next; however, after several weeks (as shown in the left panel) it falls behind that of the control locations. At the same time, revenue per transaction tends to increase. One explanation for this observation—which would be at odds with our findings from the field experiment—is that customers may react on the extensive margin; in other words they pay positive amounts in response to the appeal even though they paid zero prior to seeing the appeal. Alternatively, the response may occur on the intensive margin: there may be partially honest customers who now, following the appeal, increase their payment (or dispense with the paper) rather than be dishonest.[21]

Online Appendix A contains regressions by which one can examine the significance levels of these results and whether or not they hold when controls are included. Here, we only report some simple statistical tests based on data aggregated over the study's three periods. The average revenue per transaction is €0.0454 in period 1 (days −21 to −7), €0.0501 in period 2 (days 0 to 9), and €0.0618 in period 3 (days 14 to 23) in the treatment locations; in the control locations, the respective numbers are €0.0461 for period 1, €0.0442 for period 2, and €0.0467 for period 3. Although the DID effects support the appeal to honesty, they are not significant between periods

---

20.    A closer examination reveals that revenue increases two days after the treatment is initiated. A plausible explanation is that customers may be more attentive to a new message (this was the case in our field experiment) than to changes in an old message.

21.    Call $WTP\_N_i$ a customer's willingness to pay for obeying the honesty norm and $WTP\_I_i$ this person's willingness to pay for the paper (i.e., the value of information). Then a customer will take the paper if $WTP\_I_i \geq WTP\_N_i$. Following the intuition sketched in Figure 4, the customer will pay the full price ($P$) of the paper if $WTP\_N_i \geq P$ and otherwise will make a payment equal to $WTP\_N_i$. So if an intervention increases $WTP\_N_i$ but leaves $WTP\_I_i$ unchanged, then some former customers will drop out of the market, and new customers will enter only if they are sufficiently honest. Hence, sales will decrease and the payments that are made will come from more honest customers.

1 and 2 ($p > 0.1$, one-tailed $t$-test). However, they are (weakly) significant between periods 1 and 3 ($p < 0.1$). There is no difference again between periods 2 and 3 ($p > 0.1$).[22] At the treatment locations, there were 2,276, 1,675, and 1,681 transactions for periods 1, 2, and 3, respectively; these numbers are 2,753, 2,209, and 2,391 at the control locations for those respective periods. The DID effects are insignificant between periods 1 and 2 ($p > 0.1$, one-tailed $t$-test), but they are significant between periods 1 and 3 and also between periods 2 and 3 (each $p < 0.1$). The regressions in the Online Appendix confirm significant DID effects in periods 2 and 3 for revenue *per location*. The results are weaker for revenue *per transaction* but remain consistent with the results of the aforementioned tests. Such discrepancy is not surprising in light of the wide variation in revenue per location and in the number of transactions per location.[23]

In short, the results of this quasi–experimental study are auxiliary to those of our main field experiment. Our regressions impose strong assumptions on the standard errors, and we do not wish to overinterpret the results. Nevertheless, they lend tentative support to an interpretation according to which the appeal has a sustained positive effect in encouraging the same people to be more honest.

### 3.3. Survey Results

In this part of our study, the payments of 402 customers were collected at 43 different locations.[24] Overall, 39% of customers pay zero, 42% make a positive payment that is below the price of the paper, and 19% pay the full price. The average payment is €0.22, which is considerably higher than the average payment of €0.08 in the field experiment (where we also observed individual payment data).[25] Of the 402 customers, 215 (53%) agreed to participate in the interview. Customers who participated in the interview had made higher payments for the paper than those who declined to be interviewed (€0.258 versus €0.179, $p < 0.01$). One of the many reasons explaining this effect might be that a preference for honesty is a positive correlate of personality traits that

---

22. Because of the small sample, we do not use locations as units of observation to test for differences in average revenue per transaction. Instead, we randomly match between 15 and 16 locations into constant groups within a route. Average revenue per transaction is then calculated as revenue divided by the number of transactions for all locations within a group. Finally, our test compares the DIDs of average revenue per transaction among eight groups in the treatment locations to those groups in the control locations (see Online Appendix A for details).

23. For example, in the pre-experimental period the standard deviation of revenue per location and per day is 30.55 with a mean of 13.65; for number of transactions, the standard deviation is 1.71 with a mean of 2.98. Hence the coefficient of correlation between revenue and transactions per location is surprisingly low ($r = 0.146$).

24. The survey study was conducted independently of the field experiment and the quasi–field experiment (see Online Appendix B for details). We collected the data in four towns in Vorarlberg on three Sunday mornings in May and June 2004. The total population of these four towns was 118,500 at that time.

25. There are at least two explanations for this observation. First, the survey was conducted on Sundays whereas the experiment was run during the week. On weekdays, the paper contains less information and subscribers receive it via the postal service. Second, in the field experiment we put only one paper into a bag, whereas bags at the sales booths were stocked at their usual capacity in the survey.

affect an individual's willingness to participate in an interview. Since we collected no information on those who declined to participate in the interview, we cannot directly address the issue of selection. The survey contains a rich set of questions on social and risk-related behaviors as well as on the sociodemographic background of respondents. We hope that these variables enable a substantial reduction in the risk of bias (due to omitted variables) in the analysis that follows.

We apply a two-part hurdle model (Cragg 1971) to analyze our data. This method explicitly accounts for the possibility of separating the decision to pay from the decision of how much to pay. First, we use a probit model to estimate the likelihood of positive payments; second, we use truncated regressions to analyze the amount paid by those who do make some positive payment. The model allows us to identify variables that are relevant to whether a customer has internalized the honesty norm and—once that hurdle is crossed—her willingness to pay and thus follow the norm.[26] Estimation results are reported in Table 2.[27] In the estimates, we cluster by location of the sales booths.

Our results point to the relevance of sociodemographics. Conditional on paying, males pay €0.077 less than females (see the variable *Male* in column 6).[28] However, the same variable has no influence on the likelihood of positive payments in stage 1. People living with partners (*Partner*) pay €0.0903 more than singles. Four factors play no significant role in either stage of the estimation: the age of respondents (*Age > 50*), net household income (*Income*), degree of higher education (*High_Education*), and whether the respondent had children (*Children*).

Honesty is also associated with various aspects of social behavior. With the possible exception of the variable *Church*—which indicates whether or not the respondent regularly attends church services[29]—the correlations have their expected signs. Honesty is positively associated with people's trust in the legal system (*Trust_Legal*)

---

26.    As an alternative, we estimated a left-censored Tobit model with 0 as the cutoff point. A likelihood ratio test rejected the more restrictive Tobit model (the same coefficients explain both the censoring mechanism and the outcome) in favor of the two-part hurdle model. The test is based on the $\chi^2$ statistic with the number of independent variables (including the constant) degrees of freedom: $\chi^2(df) = 2(\ln L_{probit} + \ln L_{trunc} - \ln L_{Tobit})$. The Tobit model is rejected because the test statistic $\chi^2(169) = 265.8$ exceeds the $\chi^2$ critical values.

27.    Online Appendix B includes a detailed description of the set of independent variables along with their means and standard deviations.

28.    The marginal effects presented in this section are based on either the probit model (stage 1) or the truncated regression model (stage 2). They are calculated at sample means and represent a change in the average likelihood of paying some positive amount, a change in payment for a unit change in an independent continuous variable, or a change in a binary variable from 0 to 1.

29.    We did not ask people to state their religious denomination because the vast majority of people living in the western provinces of Austria are Roman Catholic. We can only speculate with regard to the reasons behind the large negative effect of this variable. One reason might be that church attendees lacked the coin money needed to make the payment. Active religious participation is high in the region and it is plausible that, on a typical Sunday morning, many people had already donated some of their coinage to the church. Of course, this explanation does not change the fact that, among those who pay for the newspaper, church attendees are particularly dishonest (i.e., pay much less than others).

TABLE 2. Hurdle model regressions (clustered for sales-booth locations).

|  | Probit | | | Truncated regression[a] | | |
|---|---|---|---|---|---|---|
|  | 1<br>Coef. | 2<br>S.E. | 3<br>Mfx[b] | 4<br>Coef. | 5<br>S.E. | 6<br>Mfx[c] |
| *Male* | −0.41 | 0.27 | −0.13 | −11.60** | 5.17 | −7.72 |
| *Partner* | −0.09 | 0.29 | −0.03 | 13.55* | 7.68 | 9.02 |
| *Only_Child* | 0.50* | 0.28 | 0.15 | −2.65 | 8.75 | −1.77 |
| *Age > 50* | −0.37 | 0.31 | −0.13 | −2.06 | 8.35 | −1.37 |
| *Income* | 0.15 | 0.13 | 0.05 | 1.75 | 2.55 | 1.17 |
| *High_Education* | −0.05 | 0.28 | −0.09 | 0.10 | 7.50 | 0.06 |
| *Children* | 0.35 | 0.35 | 0.12 | −0.89 | 9.98 | −0.59 |
| *Church* | −0.31 | 0.29 | −0.11 | −27.71*** | 7.61 | −18.45 |
| *Reciprocity_Pos* | 0.05 | 0.64 | 0.02 | −5.46 | 15.100 | −3.64 |
| *Reciprocity_Neg* | 0.27 | 0.25 | 0.09 | 2.86 | 6.34 | 1.90 |
| *Donate_Charity* | 0.60** | 0.28 | 0.19 | 2.11 | 9.61 | 1.41 |
| *Trust_Legal* | 0.57*** | 0.22 | 0.20 | −4.11 | 7.82 | −2.74 |
| *Volunteer* | 0.47 | 0.31 | 0.16 | 12.18** | 6.24 | 8.11 |
| *Esteem* | 0.32* | 0.17 | 0.11 | 13.83** | 5.82 | 9.20 |
| *Cheat_Tax* | −0.37 | 0.23 | −0.13 | −20.40*** | 8.06 | −13.58 |
| *Gambling* | −0.18 | 0.29 | −0.06 | 9.19 | 6.84 | 6.12 |
| *Invest* | −0.34 | 0.21 | −0.12 | 5.93 | 7.05 | 3.95 |
| *Risky_Sport* | −0.17 | 0.24 | −0.06 | 9.68 | 7.99 | 6.44 |
| *Smoker* | 0.34 | 0.25 | 0.11 | 6.08 | 6.62 | 4.05 |
| *Constant* | −0.21 | 0.67 |  | 21.09 | 18.73 |  |
| Pseudo-$R^2$ |  | 0.21 |  |  |  |  |
| Observations |  | 189[d] |  |  | 127 |  |
| $\chi^2$/*p*-value |  | 138.41/<0.001 |  |  | 81.20/<0.001 |  |

[a]Dependent variable: payment in eurocents.
[b]The marginal effect is given by $\partial E[\text{probpayment} > 0|\bar{X}]/\partial x_j$.
[c]The "conditional" marginal effect is given by $\partial E[\text{payment}|\text{payment} > 0, \bar{X}]/\partial x_j$.
[d]The number of observations in the regression is reduced from 215 to 189 because we did not have income data for some customers.
[e]Significant at 10%; **significant at 5%; ***significant at 1%.

and time spent on volunteer work (*Volunteer*).[30] The field behavior also correlates with a survey measure of tax compliance (*Cheat_Tax*) and with esteem (*Esteem*)—that is, the importance people attach to what others think of them. The significance of this last variable in both stages of the estimation may indicate that an internalized social norm complements both guilt aversion and a preference for being held in esteem by others (see Bénabou and Tirole 2006).[31] Finally, none of the variables that measure attitudes toward risk (*Gambling*, *Invest*, and *Risky_Sport*) have any statistically significant effect.

---

30.    The reader might notice an inconsistency between the effects of *Trust_Legal* and the Legal treatment in the field experiment. A possible explanation is that people with faith in the legal system are already reluctant to breach the law, and the message in the Legal treatment is insufficient to induce any changes in customers' perceptions of enforcement.

31.    For example, Gerber, Green, and Larimer (2008) report evidence suggesting that intrinsic motives interact with external forces to foster civic duty in the context of voting.

## 4. Summary and Discussion

In this study, we observe payments to an "honor system" box for newspapers under controlled field conditions. We find that customers are partially honest and that appealing to honesty increases payments on the intensive margin, although there is no movement on the extensive margin, i.e., the decision of whether a customer pays. Thus, reminding customers of legal enforcement has no effect. Suggestive evidence supports our assertions that the effect of a moral-norm reminder does not dissipate when it is posted for longer periods and that there are no negative effects when the reminder is removed.

We conjecture that the honesty reminder is effective because it reduces the moral "wiggle room" for self-deception. However, the message may also alter behavior through its effect on beliefs (see DellaVigna and Gentzkow 2010). According to a model of social norms in which a person's compliance with norms is related to the compliance of others, the reminder may shift individual reaction functions and eventually the overall outcome from a low- to a high-compliance equilibrium (see Gneezy and Rustichini 2000; Fehr and Falk 2002). If customers view their payments as being part of a strategic choice, then psychological framing effects may also interact with strategic considerations in the context of our study. Note, however, that (i) we use a nonprice manipulation which is presumably less capable of stabilizing an equilibrium; and (ii) our results are thus not consistent with a high-compliance equilibrium. In accordance with the results of laboratory research on the preference for honesty, we believe that strategic considerations are not sufficient to account for our results.

We believe that our study has important implications beyond the particular scenario under investigation. We have shown that an appeal to honesty is effective only when honesty is appropriate from the recipient's normative perspective. This finding is in line with the results of social psychology studies (see, e.g., Goldstein, Cialdini, and Griskevicius 2008) and is especially relevant in compliance research. For example, Fellner, Sausgruber, and Traxler (2013) study the effects of messages designed to increase compliance with the payment of TV license fees. They find positive effects associated with a message that highlights possible legal sanctions but find no effects for a moral appeal. The most important difference between their study and ours is that their messages are sent to a selected sample of individuals who have already chosen to deviate from the law. It is probable that these individuals belong to a group of opportunistic people who respond more strongly to external than internal incentives.

## References

Andreoni, James and Douglas B. Bernheim (2009). "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77, 1607–1636.

Barr, Abigail and Danila Serra (2010). "Corruption and Culture: An Experimental Analysis." *Journal of Public Economics*, 94, 862–869.

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts (2006). "Cues of Being Watched Enhance Cooperation in a Real-World Setting." *Biology Letters*, 2, 412–414.

Bénabou, Roland and Jean Tirole (2006). "Incentives and Prosocial Behavior." *American Economic Review*, 96, 1652–1678.

Bolton, Gary E., Elena Katok, and Rami Zwick (1998). "Dictator Game Giving: Rules of Fairness Versus Acts of Kindness." *International Journal of Game Theory*, 27, 269–299.

Bradshaw, Katherine (2003). "Coin-Operated Machines Change Single Copy Sales." *Newspaper Research Journal*, 24, 37–47.

Brandts, Jordi and Gary Charness (2003). "Truth or Consequences: An Experiment." *Management Science*, 49, 116–130.

Card, David, Stefano DellaVigna, and Ulrike Malmendier (2011). "The Role of Theory in Field Experiments." *Journal of Economic Perspectives*, 25, 39–62.

Charness, Gary and Martin Dufwenberg (2006). "Promises and Partnership." *Econometrica*, 74, 1579–1601.

Cialdini Robert B. and Noah J. Goldstein (2004). "Social Influence: Compliance and Conformity." *Annual Review of Psychology*, 55, 591–621.

Cooter, Robert (1998). "Expressive Law and Economics." *Journal of Legal Studies*, 27, 585–608.

Cragg, John G. (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica*, 39, 829–844.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33, 67–80.

Dawes, Robyn M. and Richard H. Thaler (1988). "Anomalies: Cooperation." *Journal of Economic Perspectives*, 2, 187–197.

DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence." *Annual Review of Economics*, 2, 643–669.

Ellingsen, Tore and Magnus Johannesson (2008). "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review*, 983, 990–1008.

Erard, Brian and Jonathan S. Feinstein (1994). "Honesty and Evasion in the Tax Compliance Game." *RAND Journal of Economics*, 25, 1–19.

Fehr, Ernst and Armin Falk (2002). "Psychological Foundations of Incentives." *European Economic Review*, 46, 687–724.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler (2013). "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association*, 11(3), 634–660.

Festinger, Leon (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fischbacher, Urs and Franziska Heusi (2013). *Lies in Disguise—an Experimental Study on Cheating*. Journal of the European Economic Association, 11(3), 525–547.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer (2008). "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102, 33–48.

Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner (2013). "Preferences for Truthfulness: Heterogeneity among and within Individuals." *American Economic Review*, 103, 532–548 .

Gneezy, Ayelet, Uri Gneezy, Leif D. Nelson, and Amber Brown (2010). "Shared Social Responsibility: A Field Experiment in Pay-What-You-Want Pricing and Charitable Giving." *Science*, 329, 325–327.

Gneezy, Uri (2005). "Deception: The Role of Consequences." *American Economic Review*, 95, 384–394.

Gneezy, Uri and Aldo Rustichini (2000). "A Fine is a Price." *Journal of Legal Studies*, 29, 1–18.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008). "A Room with a Viewpoint. Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research*, 35, 472–482.

Greenberg, Jerald (1990). "Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts." *Journal of Applied Psychology*, 75, 561–568.

Haan, Marco and Peter Kooreman (2002). "Free Riding and the Provision of Candy Bars." *Journal of Public Economics*, 83, 277–291.

Houser, Daniel, Stefan Vetter, and Joachim K. Winter (2012). "Fairness and cheating." *European Economic Review*, 56, 1645–1655.

Konow, James (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 90, 1072–1091.

Levitt, Steven D. (2006). "White-Collar Crime Writ Small: A Case Study of Bagels, Donuts, and the Honor System." *American Economic Review*, 96, 290–294.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research*, 45, 633–644.

Mazar, Nina and Dan Ariely (2006). "Dishonesty in Everyday Life and Its Policy Implications." *Journal of Public Policy and Marketing*, 25, 117–126.

Mittendorf, Brian (2006). "Capital Budgeting When Managers Value Both Honesty and Perquisites." *Journal of Management Accounting Research*, 18, 77–95.

Picard, Pierre (1996). "Auditing Claims in the Insurance Market with Fraud: The Credibility Issue." *Journal of Public Economics*, 63, 27–56.

Picard, Robert G. and Stephen Lacy (1999). "Legal and Economic Aspects in Theft of Newspapers: Using a Model of Newspaper Value." *Journal of Media Economics*, 12, 247–263.

Rose-Ackerman, Susan (2001). "Trust, Honesty, and Corruption: Reflections on the State-Building Process." *Archives of European Sociology*, 42, 526–570.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius (2007). "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science*, 18, 429–434.

Vanberg, Christoph (2008). "Why Do People Keep Their Promises? An Experimental Test of Two Explanations." *Econometrica*, 76, 1467–1480.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Appendix A:** The Quasi-Field Experiment—Regressions.
**Appendix B:** Details Regarding Survey-Data Collection.
**Table A.1:** OLS regressions.
**Table A.2:** OLS regressions.
**Table B.1:** Description of variables in the survey study.