This article was downloaded by: [128.76.244.219] On: 30 January 2022, At: 03:40
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation

Zhixin Dai, Fabio Galeotti, Marie Claire Villeval

To cite this article:
Zhixin Dai, Fabio Galeotti, Marie Claire Villeval (2018) Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. Management Science 64(3):1081-1100. https://doi.org/10.1287/mnsc.2016.2616

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MANAGEMENT SCIENCE
Vol. 64, No. 3, March 2018, pp. 1081–1100
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

http://pubsonline.informs.org/journal/mnsc/

Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation

Zhixin Dai,[a,b] **Fabio Galeotti,**[b] **Marie Claire Villeval**[b,c,d]

[a] China Financial Policy Research Center, School of Finance, Renmin University of China, Beijing 100872, China; [b] Univ Lyon, CNRS, GATE, Ecully F-69130, France; [c] IZA, 53113 Bonn, Germany; [d] Department of Public Finance, University of Innsbruck, 6020 Innsbruck, Austria

Contact: dzxruc@ruc.edu.cn (ZD); galeotti@gate.cnrs.fr (FG); villeval@gate.cnrs.fr (MCV)

Received: February 18, 2016
Revised: May 22, 2016
Accepted: July 14, 2016
Published Online in Articles in Advance: January 10, 2017

https://doi.org/10.1287/mnsc.2016.2616

Copyright: © 2017 INFORMS

Abstract. We conduct an artefactual field experiment using a diversified sample of passengers of public transportation to study attitudes toward dishonesty. We find that the diversity of behavior in terms of (dis)honesty in laboratory tasks and in the field correlate. Moreover, individuals who have just been fined in the field behave more honestly in the lab than the other fare dodgers, except when context is introduced. Overall, we show that simple tests of dishonesty in the lab can predict moral firmness in life, although fraudsters who care about social image cheat less when behavior can be verified ex post by the experimenter.

History: Accepted by Uri Gneezy, behavioral economics.

Funding: This research was supported by a grant from Keolis, a grant from the French National Research Agency (ANR) [FELIS Grant ANR-14-CE28-0010-01], and funding from the European Commission (Marie Skłodowska-Curie Individual Fellowship) [661645-IDEA-MSCA-IF-EF-ST]. This research was performed within the framework of the LabEx CORTEX [ANR-11-LABX-0042] of the Université de Lyon, within the program Investissements d'Avenir [ANR-11-IDEX-007] operated by the French National Research Agency.

Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2016.2616.

Keywords: dishonesty • fare dodging • field experiment • external validity • public transportation

1. Introduction

Fraud is a major concern in many spheres of economic life, from the financial sector to public transport, from sports to politics and academia, and it is responsible of enormous costs for society. For example, the overall gross tax gap in the United States represents more than 16% of the estimated actual tax liability for the years 2008–2010 (IRS 2016), and insurance fraud is estimated to be around 10% of all claim expenditures in the United States (Mazar et al. 2008). The recent Panama Papers scandal has focused the attention on the importance of evasion in tax heavens. Corporate scandals (e.g., Enron, Société Générale, Volkswagen) entail immediate repercussions on the stock market and may durably undermine consumers' trust. Understanding better the determinants of the attitudes toward fraud is undoubtedly a major challenge to develop more effective policies.

The lack of reliable data on activities that are, by definition, secretive contributes to explain the recent booming of laboratory experiments studying dishonesty.[1] These experiments have renewed considerably our knowledge by challenging the standard economics-of-crime approach (Becker 1968). In particular, they have revealed a large diversity in the attitudes toward cheating, showing that the decision to engage or not in fraud does not result only from an evaluation of benefits and costs (e.g., Erat and Gneezy 2012, Fischbacher and Föllmi-Heusi 2013, Gibson et al. 2013, Abeler et al. 2014).[2] However, although the advantages of using the laboratory to study dishonesty are many, one potential limitation arises from the fact that this setting is highly artificial and the results may not generalize to real settings. This issue, i.e., the generalizability of lab evidence, is a major source of debates and the discussion is still ongoing in the profession (see, e.g., Levitt and List 2007, Falk and Heckman 2009, Camerer 2015, Fréchette and Schotter 2015, Charness and Fehr 2015).

The main aim of this study is to analyze whether the diversity of attitudes toward cheating that has been identified by means of various laboratory tasks is correlated with the dishonesty of individuals in real life. We study to what extent the findings of laboratory experiments on dishonesty are externally valid with an ordinary population. To achieve this aim, we compare the cheating behavior of similar people in the field and in the laboratory by means of an artefactual field experiment (in the sense of Harrison and List 2004) using a diversified sample of 279 passengers of a public transport service in France.

Levitt and List (2007) identify five main threats to generalizability in a typical laboratory experiment due to the fact that decision making involves economic but also moral and ethical considerations: the scrutiny of actions by others, a lack of anonymity, the context in which the decision is embedded, the self-selection of participants, and the stakes of the game. When studying dishonesty, some of these threats may be exacerbated, making generalizability even more challenging. In particular, the fear of scrutiny may induce subjects to resist the temptation of cheating because observability increases the moral and ethical costs of cheating. Dishonesty also possesses different facets and is likely to be context dependent; thus, decisions may differ if they are taken in or outside the lab. This leaves open the question of whether economists are able, by means of a lab experiments, to make general claims about dishonesty which hold across different contexts. Moreover, if more prosocial individuals tend to self-select in participating in experiments, and if social preferences are correlated with a higher moral firmness, lab experiments may underestimate dishonesty or overestimate the impact of interventions. Although Falk and Heckman (2009) and Camerer (2015) address very thoughtfully the limitations to the generalizability of such critiques, our study has been designed such as to avoid each of these potential issues.

We chose to study dishonesty in public transportation because fare evasion is a nontrivial offence,[3] is widespread in all countries, and costs billions of dollars worldwide (see, e.g., Donohue 2010, Fürst 2015, Cour des Comptes 2016), but is surprisingly understudied.[4] In a "policy view" (in the terms of Camerer 2015), it is crucial to determine whether the lab can be convincingly used to investigate the determinants of fare evasion in the field and how it reacts to various interventions. Our study focuses on small-scale dishonesty that is more prevalent in day-to-day life with dramatic consequences for society (Mazar and Ariely 2006), and, therefore, globally relevant from an economic point of view. The level of stakes associated with fare evasion at the individual level is comparable with the low stakes in the games used in the lab, which reduces the risk of making inaccurate inference from the experimental findings due to too different levels of stakes.

Another advantage in terms of generalizability of studying fare dodging is the opportunity to observe the dishonesty of ordinary people and the accessibility of the targeted population. Indeed, a very large fraction of the population uses public transports in France, with a huge variety in terms of age, income level, occupation, and residential location. This contrasts with other recent studies of dishonesty in the lab and in the field that target much more specific populations (open-air sellers in List 2009; bankers in Cohn et al. 2014; prisoners in Cohn et al. 2015; graders in Armantier and Boly 2013; nurses in Hanna and Wang 2015). Regarding the risk of self-selection, comparisons with data collected by the operator (Keolis) indicate that our sample is relatively similar in terms of individual characteristics and that our proportion of fraudsters is not below the estimates of the company.

Our methodology combines standard laboratory and field measures of dishonesty to test whether our measures of dishonesty generalize across contexts. Precisely, studying fare evasion allowed us to collect three field measures. The first two are based on whether subjects could at the end of the experiment display a validated ticket in exchange for a new one (similar to Bucciol et al. 2013) and on self-reports.[5] For the third measure we recruited a subsample of passengers just after they paid a fine. The core of our paper consists of correlating these measures of dishonesty in the field to dishonesty in the lab, as measured in both an abstract task, a *die-under-cup task* inspired by Fischbacher and Föllmi-Heusi (2013) and Shalvi et al. (2011) that eliminates any possible scrutiny,[6] and a contextualized task, a *public transportation game* where subjects decide to buy or not a ticket to use a fictional bus. We introduced this public transportation game because it allows us to measure cheating behavior at the individual level, whereas the die task only allows us to compare the distribution of all subjects' reports to the theoretical distribution of random draws. Note that behavior may differ in this game compared to the die task for at least two psychological reasons. Since the experimenter is able to identify ex post in the data who has cheated in this game, a feeling of scrutiny may affect negatively the decision to lie (if subjects experience a disutility from being exposed as a cheater in the experimenter's eyes). On the other hand, introducing a public transportation context may prime the identity of passengers, which may reinforce the fare dodgers' willingness to cheat. Our objective, however, is not to compare cheating in the two tasks, but to test whether the correlation between behavior in the field and in the lab holds in these two different environments.

As a robustness test, we conducted a complementary study with a standard student-subject pool.[7] To get a field measure of dishonesty, we overpaid these subjects at the end of the session and checked whether they reported the error or not (like Gneezy et al. 2014, Potters and Stoop 2016).

Our contribution and novelty are fourfold. First, by presenting the first study comparing the behavior of actual fare dodgers and non–fare dodgers across experimental tasks, we are able to investigate whether or not the same persons cheat in different contexts and to validate the external validity of lab measures of dishonesty with and without context.[8] This constitutes a test of the generalizability of lab findings on dishonesty to the field that does not rely on priming methods (such as Cohn et al. 2014, 2015) but on direct observation, and that involves ordinary people. Second, we can study the consistency of objective measures of cheating with self-reports and whether lab measures correlate more with objective measures or with self-reports. Deviations allow us to identify image concerns. Third, we provide the first study able to measure the immediate impact of a fine in the field on subsequent dishonesty in the lab, by testing whether fare dodgers who have just paid a fine fraud more or less than others when given a new opportunity to cheat. Fourth, we contribute to a better knowledge of fare dodgers, their moral norms, and their individual characteristics.

Our results show that the lab measures of dishonesty correlate with rule violation in real life. The proportion of fully dishonest subjects in the die task is higher among passengers who do not hold a validated ticket, while the proportion of incomplete liars is the highest among those who hold a valid ticket but self-report as occasional fare dodgers. Our study with students confirms the correlation between field and lab measures of dishonesty. Moreover, fare dodgers who had just paid a fine behave more honestly than the other fare dodgers in the die task but not when context is introduced, which suggests a priming effect of context. Behavior in the contextualized task is essentially consistent with behavior in the die task and with our field measures of dishonesty, except for passengers without a ticket who, however, self-report as nonfraudsters. Indeed, when decisions can be verified ex post, self-image concerned people cheat less than when scrutiny is impossible. Overall, we show that simple tests of dishonesty in the lab are able to capture individual attributes that predict rule violation by ordinary people in real life. Showing the external validity of the die-under-cup task is a major finding given the large interest for this experimental paradigm in the recent literature on dishonesty. More generally, this constitutes a major step in the validation of the use of the laboratory to study how dishonesty reacts to policy interventions. Caution must, however, be taken when behavior can be verified by the experimenter since it may underestimate the dishonesty of image-concerned individuals.

The remainder of this paper is as follows. Section 2 reviews the related literature. Section 3 describes the experimental design and procedures. Section 4 details our results, and Section 5 presents our replication study. Section 6 discusses the results and concludes.

2. Related Literature

This section briefly reviews the various methods used by previous experimental studies to investigate the link between dishonesty in the field and in the laboratory and characterizes how we differ from these studies.

One approach consists of examining how societal norms across countries shape individuals' intrinsic standards of honesty and ethical values, as measured in the lab. With a lab experiment conducted in 23 countries, Gächter and Schultz (2016) have demonstrated that individual dishonesty is directly influenced by the prevalence of rule violations (such as tax evasion, political fraud and corruption) in the society. In contrast, our study examines the correlation between the behavior of the same persons in the field and in the lab in a given culture.

Another approach consists of priming the identity of people with a presumed or proven culture of cheating. Cohn et al. (2014) show that bank employees, when their professional identity is primed, behave more dishonestly in a coin-tossing task compared to control groups. This study demonstrates the impact of business culture but cannot relate directly (unobserved) individual dishonesty in the field to that observed in the lab. Cohn et al. (2015) identify a causal effect of criminal identity on dishonesty. Indeed, inmates behave more dishonestly when their criminal identity is primed compared to regular citizens, and the dishonesty of primed and nonprimed prisoners, pooled together, is correlated to the offences committed in prison. This provides evidence of the external validity of a coin-tossing task, but with a very peculiar population characterized by persons who committed serious crimes. In addition, a large proportion of the inmates were convicted for crimes that are not necessarily linked to dishonesty. In contrast, our study focuses on the day-to-day dishonesty of ordinary people, which involves small-scale cheating, and we replace priming with a direct observation of individual dishonesty in a real setting.[9]

A different approach consists of replicating the same experiment in field and lab settings. Conducting a field experiment and two lab experiments on bribing, Armantier and Boly (2013) obtain analogous treatment effects, both in terms of direction and magnitude. However, since bribing involves *joint* dishonesty, it may be partly motivated by social preferences and what works for corruption may not work for other forms of dishonesty. More importantly, while the authors uncover a clear parallelism between the lab and the field, they do not test whether corrupt individuals in the lab would adopt the same behavior in the real world. The same applies to Alm et al. (2015), who show the similarity of tax evasion decisions of actual taxpayers, students, and nonstudent subjects in comparable field and lab settings, and to Choo et al. (2014), who find a lower compliance of students. By measuring the behavior of identical individuals in the field and in the lab, we believe that we provide a stronger test of the generalizability of lab experiments on dishonesty.

Indeed, another method consists of observing the dishonest behavior of identical individuals both in a naturally occurring environment and in the lab. A secondary result of a study by List (2009) on open-air markets is that cheating, intended as maintaining collusive promises between sellers, correlates between a natural and a framed field experiment, whereas little correlation is found between cheating rates in the lab and in the other environments when no context is introduced. This last result is, however, based on a small sample of 17 dishonest sellers. In the spirit of List (2009), we compare the behavior of identical persons in a naturally occurring situation and in the lab, we manipulate the context of the games, and we conduct a similar study with student subjects. In contrast, our subjects are aware of participating in an experiment. Our approach, based on a larger and more general population and on a standard lying task, can be very useful to complement previous research and provide substantial evidence on the external validity of lab findings.[10]

The experiment most closely related to our study is that of Hanna and Wang (2015), run in India. It shows that nurses who report more favorable outcomes in a die-rolling task are more likely to be absent from work.[11] Like us, the authors aim to validate a laboratory measure of dishonesty using a nonstudent subject pool by correlating it with a field measure of dishonesty—fraudulent absenteeism. In contrast, our study involves a more generic sample of the general population, uses monetary incentives, compares different measures of dishonesty in the field, and test the predictive value of alternative lab measures with and without context. It thus provides a more complete framework to understand whether dishonesty can be reliably measured in the lab.[12]

3. Experimental Procedures and Design

In this section, we first present our recruitment methods and experimental procedures. Then, we describe the content of the experiment. Next, we explain our measures of dishonesty in the laboratory and in the field. Finally, we describe our sample of subjects.

3.1. Recruitment and Procedures

The experiment was conducted in Lyon (France) during nine consecutive days (excluding Sundays) from 11 A.M to 7 P.M. Recruiters wore a badge (with the word “survey” (“*enquête*”) written on it) to prevent being perceived as inspectors. They were instructed to invite subjects at the precise arrival of trams and buses at the Gare Part-Dieu stop, located in front of the biggest shopping mall in the city center and the main train station. The procedure consisted of randomly approaching passengers who were getting off the vehicle by asking them whether they were holding a monthly pass. The invitation to participate was extended only to passengers without a pass to maximize the number of potential fare dodgers participating in the experiment.[13] The experiment was presented as a study ran by researchers from the university about means of transportation. Potential participants were told that the duration of the study was about 45 minutes, that they would have to answer questions, that the study was anonymous and did not require any particular knowledge, and that a minimum of €25 would be paid in gift vouchers immediately at the end of the study. The recruiters were not informed about the aim of the study, and if someone asked about its content, they were instructed to tell them that they did not know. Once a passenger agreed to participate, he was asked to keep his used ticket if he had one because it would be replaced with a new one at the end of the experiment.

The 244 subjects who agreed to participate were escorted to a building located 60 meters from the stop, where we had installed our mobile lab (see Online Appendix 1). Upon arrival, each subject signed a consent form, received a tablet, and was assigned to a desk. Each desk was isolated by means of mobile partitions. Since the experiment was based on individual decisions, we had a permanent inflow of participants with a maximum of nine subjects at the same time. We answered questions in private. Once a subject completed the experiment, he was escorted to a separate room and paid out in gift vouchers. After giving the vouchers, the assistant asked the subject whether he had a validated ticket in the purpose of exchanging it with a free day pass since the ticket was no longer valid.

To measure the immediate impact of a sanction on honesty, we also recruited 35 subjects from the Fine Collection Office (FCO), where passengers who are caught to travel without a valid ticket and do not pay their fine on the spot can pay their fine. These subjects were invited immediately after paying their fine. The rest of the procedure was similar to that used with the other passengers. The experimenter knew when a person was coming from the FCO because he was escorted by different recruiters than those who were at the bus and tram stop. This information was recorded by a code entered in the tablet.

A total of 279 subjects participated. To get a sample of passengers as representative as possible, participants were sampled based on a stratified random procedure, which followed a compulsory survey conducted by the local public transport agency TCL-Sytral in Lyon in 2014. Online Appendix 2 compares our subject pool with the population of this representative survey. Our sample differs mainly in terms of age (we excluded passengers younger than 18, we slightly overrepresent passengers from 18 to 24 years old, and we underrepresent passengers older than 60) and status (we overrepresent unemployed people and underrepresent white collars and retirees).

3.2. Content of the Experiment

The experiment consists of four parts and a final questionnaire. It was fully computerized, including the instructions that were displayed on the tablet at the beginning of each part (see Online Appendix 3). For the purpose of this paper, we focus only on the first three parts.[14]

3.2.1. First Part: Elicitation of Attitudes Toward Risk, Ambiguity, and Losses. We used Eckel and Grossman's (2008) method with the representation proposed by Eckel et al. (2012). Precisely, subjects have to choose one lottery in each of four successive sets of six binary lotteries displayed on a ring in different screens. The expected payoff of each lottery increases as one moves clockwise, from €28 to €36, as well as the variance of payoffs. The first set of lotteries is used to elicit risk attitudes in the gain domain. Each lottery offers two outcomes, each with an equal likelihood. The second set offers the same choices but in the domain of losses. To hold expected earnings constant across tasks, subjects receive an initial endowment of €40. The comparison of the choices made in these two tasks gives a measure of the subjects' degree of loss aversion. The next two sets of lotteries offer the same possible outcomes in the domain of gains, first, and of losses, next. Like in Cardenas and Carpenter (2013), the probability of each outcome is, however, uncertain, with a minimum of 30% and a maximum of 70%. Comparing choices in the first and the third (second and fourth, respectively) sets gives a measure of ambiguity aversion in the gain (loss) domain.

3.2.2. Second Part: Die-Under-Cup Task. To elicit the subjects' proneness to lying, we adapted the procedure of Fischbacher and Föllmi-Heusi (2013). Subjects are first asked to fill out a brief questionnaire.[15] They are then told that, for answering these questions, they can earn €0, €3, or €5, depending on a die roll. The die is inside a sealed cup that allows only the subject to see the die, as in Shalvi et al. (2011). We use a six-faced die with three colors rather than numbers. The die can give three possible outcomes (*red, yellow, blue*) with equal probability (1/3). Subjects are instructed to privately roll the die twice and report the outcomes of the two rolls in the tablet. The first roll determines the earnings of the task: €0 if the color reported is blue, €3 if yellow, and €5 if red. The second roll does not give rise to any payment.[16] To avoid any error, the text above each reporting screen on the tablet reminds the subjects whether his report determines payment or not. This task allows us to collect information on full honesty and partial and full dishonesty at the aggregate level, without losing too much statistical power since we have only three possible outcomes rather than six, in contrast to previous die-rolling experiments.

3.2.3. Third Part: Public Transportation Game. This part departs from the abstract environment of the die task to move closer to a naturally occurring situation of public transportation. The task is depicted as a decision on whether to buy or not a ticket for using an imaginary bus. Seven destinations can be reached, each described as a subsequent stop of the bus. For every possible final destination (stops 1–7), the subject has to decide whether to buy a ticket or not. One decision, chosen at random at the end of the experiment, is relevant for payment. The price of the ticket equals €1.70 irrespective of the destination (equivalent to the actual tariff). Subjects benefit from using the bus, which is explained in terms of time saved. We used two treatments. In the "Variable treatment" (144 subjects), the benefit increases in the distance between the departure point and the final destination, in increments of €2 per stop, with an expected benefit of €8. In the "constant treatment" (135 subjects), the benefit is €8, irrespective of the destination. The variable treatment aimed to test whether lies are more likely when their relative benefit increases (i.e., when the distance is shorter) rather than when it is fixed (i.e., in the constant treatment). It turns out that there is not much of a difference between the two treatments (people cheat more on shorter distances in both treatments), and thus we pool them in the data analysis. To preserve some comparability with the die task, in period 1 subjects are informed that there is no ticket control on the line. Thus, it is clear that lying cannot be punished.[17] Since previous studies have shown that the introduction of context in a protocol can affect behavior (Cooper et al. 1999) or matters for ecological validity (Fréchette 2015), this part allows us to test whether the difference in attitudes toward dishonesty between actual fare dodgers and non–fare dodgers is larger in the contextualized game than in the abstract die task. More importantly, this task allows us to measure dishonesty at the individual level, in contrast to the die task. The counterpart is that some subjects may feel scrutinized since it is possible for the experimenter to check ex post whether they lied or not.

3.2.4. Final Questionnaire. At the end of the experiment, subjects filled in a final questionnaire where we collected information about the demographic and socioeconomic characteristics of the subjects (age, gender, income category, status, residence), their usage of public transport (frequency, time), and, finally, their beliefs and opinion about fare evasion. A full list of the items used in this questionnaire can be found in Online Appendix 3.

3.2.5. Payments. We paid the earnings of one randomly selected lottery choice in part 1, those associated to the reported outcome in part 2, and either the earnings of one decision in one period of part 3 or the earnings of three randomly selected periods in part 4. Subjects received between €25 and €110, and, on average, €60.

3.3. Measures of Dishonesty

We collected, in total, two experimental and three field measures of dishonesty. One experimental measure comes from the die task and the other one is given by the public transportation game. In part 2, we can assess the extent to which our subjects are dishonest by comparing the distribution of the reported outcomes of the die roll with the uniform distribution of a fair die roll. In part 3, we measure whether and how often subjects do not buy a ticket in the game. While the first measure cannot be evaluated for each individual but only for clusters of subjects, the second one is calculated individually.

Turning to the field measures of dishonesty, one is based on whether a subject is able to display or not a validated ticket at the end of the experiment (like in Bucciol et al. 2013). We believe that this measure is quite reliable. Subjects were, at the recruitment stage, told to keep their ticket with them for replacement with a new one. The motivation was credible since the company uses a time-based pricing system and a ticket is valid only within one hour; therefore, it makes sense to compensate a ticket holder with a new ticket for the time spent in the study.[18] Even if a passenger does not plan to continue his journey, it is attractive to get a free day pass that can be used any time in exchange for a used ticket. The advantage of this measure is that we can convincingly classify subjects who are not able to display a valid ticket as fare dodgers. One limitation, however, is that it is instantaneous; it only captures subjects who evaded the fare on that specific day.[19]

Our second field measure is based on self-reports in a final questionnaire. Subjects were asked to report how many times, on average, they travel without a ticket out of 10 trips. We classify as cheaters those who report to evade at least once out of 10 trips. The advantage of this measure is that subjects who admit to travel without a ticket can be credibly classified as fare dodgers since they have no reason to lie. The disadvantage is that image-concerned fare dodgers may not be willing to admit that they are a cheater.[20]

Finally, the subjects recruited at the FCO have, by definition, acted dishonestly at least once, since they have been caught during an inspection. We classify them as fare dodgers but we expect them to behave differently in the experiment from the other recruited fare dodgers. On the one hand, they may act more honestly. Indeed, paying a fine may have an "educative" effect on immediate behavior because these people have just been reminded that violating the rules leads to punishment. Fare dodgers who pay the fine may also, in general, be more honest than other fare dodgers (as a large fraction of detected fraudsters who do not pay on the spot never pay their fine). On the other hand, they may behave less honestly because of loss repair: they may try to recover the amount of the fine by cheating in the experiment (Kastlunger et al. 2009, Dai et al. 2015). For these reasons, these subjects are classified as standalone category and their behavior is analyzed separately.

3.4. Self-Selection of Subjects

One might be concerned that evaders are less willing to take part in a study on public transports than passengers who used to pay their fare, and that the fare dodgers who agree to participate are those who are the least embarrassed by a violation of the honesty norm (although they are not aware of the aim of the study). We explain why we believe that these two sources of selection bias have not plagued our study. Table 1 reports the distribution of the subjects based on our measures of fraud in the field.

First, Table 1 indicates a high share of fare dodgers among the participants. Excluding the FCO subjects, 41.80% of the subjects were not able to display a validated ticket[21] and 54.92% reported travelling without a ticket at least once every 10 trips. The representatives

Table 1. Distribution of Subjects by Measure of Fraud in the Field

	Number of trips without a ticket, out of 10 trips (self-reported)					
	0	1 or 2	3 or more	Total	Percentages	
Ticket holders	77 (54.23)	43 (30.28)	22 (15.49)	142 (100)	58.20	50.90
Non–ticket holders	33 (32.35)	28 (27.45)	41 (40.20)	102 (100)	41.80	36.56
Total (exc. FCO)	110 (45.08)	71 (29.10)	63 (25.82)	244 (100)	100	—
FCO	13 (37.14)	9 (25.72)	13 (37.14)	35 (100)	—	12.54
Total (all)	123 (44.09)	80 (28.67)	76 (27.24)	279 (100)	—	100

Note. Percentages are in parentheses.

of the operator confirmed that our values are consistent with their measures for passengers without a pass on the lines where we recruited our subjects.[22] This indicates that fare dodgers did not self-select out of our study.

Second, in the final questionnaire, we asked subjects to indicate how they perceive a passenger who does not pay his fare on a scale from 1, for "I very strongly disapprove," to 7, for "I very strongly approve." In response, 38.24% of the non–ticket holders and 27.56% of the self-reported fraudsters reported a value lower than 4 (more or less disapprove), while 26.47% and 25.64%, respectively, reported a value higher than 4 (more or less approve). The corresponding percentages are 49.30% and 16.90%, respectively, for the ticket holders, and 64.23% and 15.45% for the self-reported nonfraudsters. If only fare dodgers with no shame self-selected in our study, we should expect a higher approval rate among these subjects. Moreover, a fraction of the subjects unable to show us a validated ticket self-reported as nonfraudsters, which suggests that these subjects care about their image.

4. Results

In this section, we first consider the general prevalence of cheating in the die task to test whether we are able to replicate previous findings. We then divide subjects in categories based on our field measures of dishonesty to study whether behavior in the die task is related to dishonesty in the field. Next, we investigate whether the relationship between dishonesty in the field and behavior in the lab holds in the contextualized transportation game. We then test the extent to which the lab behavior of our different categories of subjects differs between the public transportation game and the die task. Finally, we report the results of our complementary study ran with students.

4.1. Diversity of Attitudes Toward Lying in the Die-Under-Cup Task

As in previous studies, we find that the distribution of the reported outcomes of the first die roll (the one that determines payoffs) is significantly different from a uniform distribution (χ^2 goodness-of-fit test, $p < 0.001$).[23] Indeed, 11.11% of the subjects report the lowest outcome (blue, which pays €0), and 51.97% report the highest outcome (red, which pays €5), both of which are significantly different from the expected value of 33% (binomial tests, $p < 0.001$); 36.92% of the subjects report the medium outcome (yellow, which pays €3), which does not differ significantly from 33% ($p = 0.205$).

Assuming that subjects never misreport an outcome when it is not beneficial for them, we can nonparametrically estimate the percentages of dishonest and honest people in our subject pool (for a derivation of these estimations, see Online Appendix 4). The estimated percentage of honest subjects is 33%, while the percentage of income maximizers—who lie by reporting the highest outcome—is 28% at maximum. These values are consistent with the estimates of Fischbacher and Föllmi-Heusi (2013), who found, respectively, 39% and 22%, using a six-sided die. We can further estimate the percentages of partially and fully dishonest subjects. Fully dishonest subjects are those who report the most favorable outcome regardless of the actual outcome. We consider two definitions of partial dishonesty. The lower bound of partial dishonesty is given by the subjects who report the medium outcome when they roll the worst one, assuming that only a fully dishonest subject reports the highest outcome when it is not true. The upper bound is that partially dishonest subjects always lie a bit by reporting the next superior outcome (yellow when getting blue, and red when getting yellow). This assumes that both fully and partially dishonest subjects may lie by reporting the highest outcome. Based on these definitions, the percentage of fully dishonest subjects lies in between 0% and 28%, and that of partially dishonest subjects between 39% and 67%.

To sum up, consistently with previous studies, we find a considerable diversity of attitudes toward lying: a nontrivial fraction of subjects are fully honest in the die task, while the majority behave dishonestly but not in a way that maximizes their earnings.

4.2. Dishonesty in the Field and Lying in the Die-Under-Cup Task

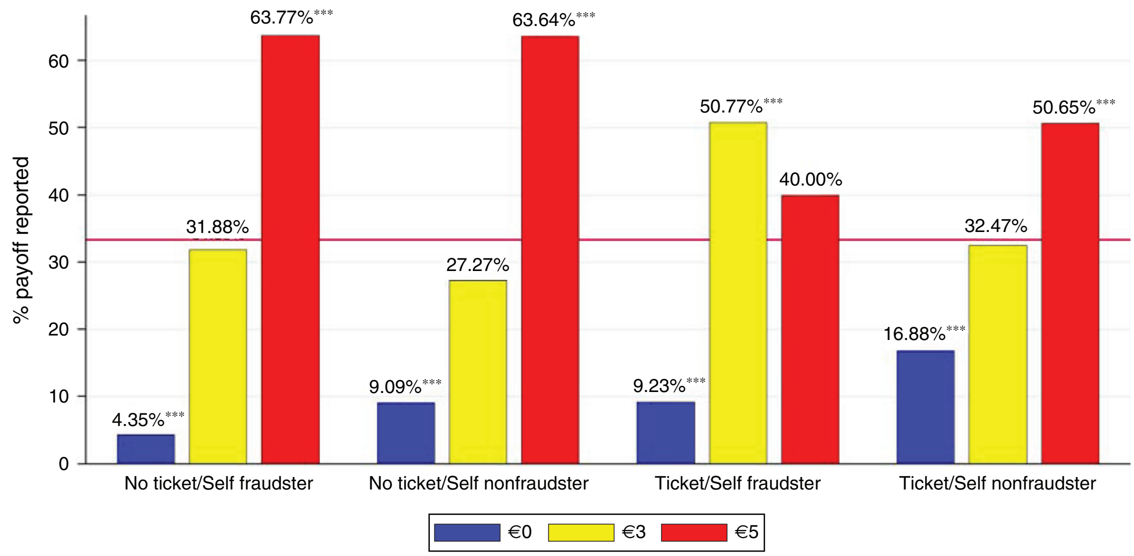
We now examine the links between dishonesty in the field and behavior in the die task. Figure 1 and Table 2 summarize the distribution of reported outcomes of the first die roll for different categories of subjects: self-reported fraudsters who do not hold a validated ticket (69 subjects), self-reported fraudsters who hold a validated ticket (65), self-reported nonfraudsters who do not hold a validated ticket (33), and self-reported nonfraudsters who hold a validated ticket (77). They also report the significance of two-sided binomial tests that the observed percentages differ from the expected 33% and Table 2 displays the p-value of a χ^2 test that the observed distribution differs from the expected uniform distribution. These statistics are used as supports for our main results.

We state our first result as follows:

Result 1. Irrespective of which field measure of dishonesty we employ, both fraudsters and nonfraudsters overreport the best outcome and underreport the worst one in the die task.

Support for Result 1. Figure 1 shows that both fraudsters and nonfraudsters tend to misreport the first die roll to their advantage. More specifically, the distributions of the outcomes reported by ticket

Figure 1. (Color online) Reported Outcomes of the First Die Roll, by Category of Subjects



Note. The horizontal line identifies the uniform distribution and is located at 33.33%.
***$p < 0.01$ (two-sided binomial tests that the observed percentage differs from 33.33%).

holders, non–ticket holders, self-reported fraudsters, and self-reported nonfraudsters differ significantly from a uniform distribution (see χ^2 tests in Table 2). For each category, the percentage of reds (blues) is significantly above (below) the expected value of 33% (see binomial tests in Table 2). The percentage of yellows is weakly significantly above 33% only for those who hold a ticket and self-report as fraudsters. The fact that ticket holders who self-report as nonfraudsters report more reds than expected by a uniform distribution is not that surprising since this category also includes a fraction of fare dodgers who lied about their status. This leads to our second result:

Result 2. (a) Fare evasion in the field is associated with more dishonesty in the die task. (b) The relationship is stronger when we consider actual rather than self-reported fare evasion.

Support for Result 2. Comparing the distributions of the outcomes reported by fraudsters and nonfraudsters in the field reveals that non–ticket holders lie significantly more than ticket holders (Fisher's exact test, $p = 0.013$), and self-reported fraudsters lie more than self-reported nonfraudsters ($p = 0.067$). In particular, non–ticket holders report the best outcome significantly more often than ticket holders ($p = 0.006$), and the worst outcome significantly less often (0.085). Self-reported fraudsters report the worst outcome significantly less often than self-reported nonfraudsters ($p = 0.056$), but report a similar proportion of reds as self-reported nonfraudsters ($p = 0.797$). Since self-reported

Table 2. Summary of the Distribution of Reported Outcomes of the First Die Roll, by Category of Subjects

Reported outcome	Blue (€0)	Yellow (€3)	Red (€5)	Avg. payoff	χ^2test
All subjects	11.11 (<0.001)	36.92 (0.205)	51.97 (<0.001)	3.71	<0.001
Fine Collection Office	17.14 (0.048)	40.00 (0.473)	42.86 (0.281)	3.34	0.124
Ticket holders	13.38 (<0.001)	40.85 (0.062)	45.77 (0.002)	3.51	<0.001
Non–ticket holders	5.88 (<0.001)	30.39 (0.600)	63.73 (<0.001)	4.10	<0.001
Self-reported nonfraudsters	14.55 (<0.001)	30.91 (0.615)	54.55 (<0.001)	3.66	<0.001
Self-reported fraudsters	6.72 (<0.001)	41.04 (0.066)	52.24 (<0.001)	3.84	<0.001
1 or 2 frauds/10 trips	9.86 (<0.001)	42.25 (0.130)	47.89 (0.012)	3.66	<0.001
3 or more frauds/10 trips	3.17 (<0.001)	39.68 (0.287)	57.14 (<0.001)	4.05	<0.001
No ticket/Self-reported fraudsters	4.35 (<0.001)	31.88 (0.899)	63.77 (<0.001)	4.15	<0.001
No ticket/Self-reported nonfraudsters	9.09 (0.002)	27.27 (0.580)	63.64 (0.001)	4.00	<0.001
Ticket/Self-reported fraudsters	9.23 (<0.001)	50.77 (0.004)	40.00 (0.292)	3.52	<0.001
Ticket/Self-reported nonfraudsters	16.88 (0.002)	32.47 (1.000)	50.65 (0.002)	3.51	0.001

Notes. The first three columns display the percentage of subjects who report the corresponding outcome and, in parentheses, the p-value of binomial tests that the percentage differs from 33.33%. The last column displays the p-value of a χ^2 goodness-of-fit test that the observed distribution differs from the uniform distribution.

fare evasion also gives an indication of the intensity of dishonesty in public transportation, we explore further the link between self-reports and dishonesty in our task. This leads to our third result:

Result 3. (a) Subjects who report cheating more frequently in the field more often misreport the outcome of the die roll, especially when they roll the least profitable outcome. (b) Occasional fare dodgers (i.e., ticket holders who confess to cheat sometimes) tend to lie partially rather than fully.

Support for Result 3. The distribution of die outcomes reported by those subjects who tell they never travel without a ticket is significantly different from that of those who confess cheating thrice or more every 10 trips (Fisher's exact test, $p = 0.047$), but not from that of those who report cheating only once or twice ($p = 0.274$). Subjects who cheat thrice or more report *blue* in only 3.17% of the cases, a percentage significantly below the proportion of blue outcomes (14.55%) reported by subjects who report they never travel without a ticket ($p = 0.020$).

As shown by Figure 1 and the χ^2 tests in Table 2, the percentage of reds is significantly above 33% for each category of subjects, except the self-reported fraudsters holding a ticket. Binomial tests in Table 2 indicate that in this latter category, subjects report significantly more yellows, and not reds, than expected, which gives evidence of less frequent full dishonesty in this group. Pairwise comparisons show that non–ticket holders display a similar lying behavior irrespectively of whether they self-report as fraudsters or nonfraudsters (Fisher's exact test, $p = 0.564$), and they in general lie more than ticket holders.[24] In contrast, among ticket holders, the distribution of reported outcomes differs significantly between self-reported fraudsters and nonfraudsters ($p = 0.077$). The self-reported fraudsters holding a ticket are probably occasional fare dodgers, as they admit to sometimes travelling without a ticket but they carry a valid one on the day of the experiment.

Finally, we consider the behavior of the subjects who have just paid a fine.

Result 4. Subjects who have just paid a fine cheat less in the die task than the other subjects who do not hold a validated ticket.

Support for Result 4. The subjects from the FCO can be classified as fare dodgers. Yet, interestingly, their behavior in the die task follows a uniform distribution (see χ^2 test in Table 2). If the percentage of blues differs significantly from the expected value of 33%, this is not the case for the percentages of reds and yellows (see binomial tests in Table 2). The distribution of outcomes reported by these subjects is similar to that of ticket holders (Fisher's exact test, $p = 0.841$), and it differs from that of non–ticket holders ($p = 0.040$). In particular, the FCO subjects report a similar proportion of reds, blues, and yellows as non–fare dodgers ($p > 0.1$), but more blues and fewer reds than other fare dodgers ($p = 0.076$ and 0.046, respectively).

Additional support for Results 2–4. Table 3 displays the percentages of honest and fully and partially dishonest subjects in each category, as estimated nonparametrically. It confirms that there are fewer fully dishonest and more honest people among ticket holders and subjects who have just paid a fine. The largest (smallest) percentage of fully dishonest (honest) people is among subjects who travelled without a ticket the day of the experiment. Subjects who self-report as fraudsters tend to be mostly partially dishonest, especially those who were able to display a validated ticket (i.e., occasional fraudsters).

Table 3. Percentages of Honest and (Partially and Fully) Dishonest Subjects (Die Task)

			Partially dishonest		Fully dishonest	
Categories of subjects in the field	Fully honest (%)	Income maximizers (%)	Lower bound (%)	Upper bound (%)	Lower bound (%)	Upper bound (%)
All	33	28	39	67	0	28
Fine Collection Office	51	14	34	49	0	14
Ticket holders	40	19	41	60	0	19
Non–ticket holders	18	46	37	74	9	46
Self-reported nonfraudsters	44	32	25	49	7	32
Self-reported fraudsters	20	28	51	80	0	28
1 or 2 frauds/10 trips	30	22	49	70	0	22
3 or more frauds/10 trips	10	36	55	90	0	36
No ticket/Self-reported fraudsters	13	46	41	83	4	46
No ticket/Self-reported nonfraudsters	27	45	27	55	18	45
Ticket/Self-reported fraudsters	28	10	62	72	0	10
Ticket/Self-reported nonfraudsters	51	26	23	47	3	26

Table 4. Determinants of the First Roll Outcome Reported in the Die Task (Multinomial Logit Regressions)

	(1)						(2)					
	Blue (€0)		Yellow (€3)		Red (€5)		Blue (€0)		Yellow (€3)		Red (€5)	
	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE
Ticket holder	0.069	0.048	0.124*	0.064	−0.193***	0.063	0.092*	0.051	0.127*	0.068	−0.218***	0.069
Self-reported nonfraudster	0.069	0.043	−0.132**	0.063	0.063	0.065	0.146**	0.059	−0.186**	0.074	0.04	0.08
Fine Collection Office	0.139**	0.063	0.051	0.094	−0.190**	0.096	0.127**	0.053	0.020	0.092	−0.146	0.094
Individual characteristics	No		No		No		Yes		Yes		Yes	
Number of observations	279						279					
Pseudo-R^2	0.031						0.145					
df	6						58					
Prob > F	0.02						0.002					

	(3)						(4)					
	Blue (€0)		Yellow (€3)		Red (€5)		Blue (€0)		Yellow (€3)		Red (€5)	
	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE	dy/dx	SE
No ticket and self-reported nonfraudster	0.077	0.082	−0.065	0.107	−0.012	0.107	0.128	0.078	−0.044	0.106	−0.085	0.11
Ticket and self-reported fraudster	0.079	0.07	0.161**	0.08	−0.240***	0.083	0.086	0.075	0.209**	0.085	−0.295***	0.092
Ticket and self-reported nonfraudster	0.142**	0.065	−0.009	0.081	−0.133	0.082	0.234***	0.068	−0.063	0.088	−0.171*	0.09
Fine Collection Office	0.144**	0.072	0.069	0.098	−0.213**	0.1	0.122**	0.059	0.055	0.094	−0.178*	0.097
Individual characteristics	No		No		No		Yes		Yes		Yes	
Number of observations	279						279					
Pseudo-R^2	0.032						0.151					
df	8						60					
Prob > F	0.043						0.001					

Notes. Multinomial logit regressions with robust standard errors are shown. The table reports the marginal effects. The "dy/dx" notation indicates the derivative, i.e., the rate of change of the variable y in reaction to a change in variable x. Independent variables include individual characteristics, attitudes toward risk, usage of public transportation, norms, and beliefs. The full regressions are contained in Online Appendix 6. In models (3) and (4), the omitted category is "No Ticket and self-reported fraudster."

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Additional support for these results comes from a regression analysis in which we control for the socioeconomic and behavioral characteristics of the subjects and their beliefs and moral norms about honesty, since they may drive the differences in behavior between fare dodgers and non–fare dodgers. Indeed, we have found that the sample of fare dodgers differs from the sample of non–fare dodgers in many respects.[25] Table 4 reports marginal effects in four multinomial logit regressions with robust standard errors in which the dependent variable is the reported outcome of the first die roll.[26] In model (1), the independent variables include a dummy for holding a ticket, one for self-reporting as a non–fare dodger, and one for being recruited at the FCO. Model (2) controls for individual characteristics (being a male, age, having a monthly income of less than €1,000, being a student, being employed, having a high school education or lower, living in the city center, living in an upper class neighborhood, and number of social networks), attitudes toward risk (risk attitude, loss aversion, and ambiguity aversion, as measured in the first part of the experiment), and usage of public transportation (frequency of trips in the previous month, whether the subject travels after 7 p.m., whether the mean duration of the trips is less than 15 minutes, and whether the subject uses the bus exclusively, the tram exclusively, or both the bus and tram). We also control for moral norms (degree of approval on fare evasion, minimum proportion of trips without a ticket to be considered a fare dodger) and beliefs (belief on the mean fraud rate in transportation, belief on the mean probability of control, whether the subject overestimates the fine, whether he underestimates it). Models (3) and (4) are similar to (1) and (2), respectively, except that we distinguish between the four categories of subjects, taking the non–ticket holders who acknowledge they are fraudsters as the omitted category.

Table 4 confirms that a subject without a ticket is more likely to report the highest die outcome, red, even after controlling for individual characteristics. A self-reported fraudster is more likely to report yellow, and less likely to report blue especially if he does not hold a ticket. Finally, a subject who has just paid a fine reports more honestly than a non–ticket holder.[27]

4.3. Dishonesty in the Field and Evasion in the Public Transportation Game

Our contextualized public transportation game confirms the general relationship between dishonesty in the lab and in the field with some differences with the die task, which we summarize in Result 5.

Result 5. Dishonesty in the public transportation game is associated with dishonesty in the field but, contrary to the die task, (a) the association is stronger when we consider self-reported dishonesty rather than holding or not a ticket; (b) among self-reported nonfraudsters, non–ticket holders behave as honestly as ticket holders; and (c) subjects who have just paid a fine do not behave more honestly than non–ticket holders.

Support for Result 5. Since we have multiple observations per subject (one for each of the seven destinations), we measure dishonesty in this game as the *rate of ticket fraud*, i.e., the proportion of times out of all destinations a subject is willing to travel without a ticket knowing that no control is exerted (we get one independent observation per subject.) Since we find no difference in fraud rates between the variable and the constant treatments (Mann–Whitney test, $p = 0.167$), we pool the data together. Table 5 displays the fraud rate and the percentages of fully honest and partially and fully dishonest subjects in the game, for the whole sample and for each field category. Subjects who always buy a ticket in the game are classified as *fully honest*, those who never buy one are classified as *fully dishonest*, and those who do not buy a ticket at least once but not systematically are classified as *partially dishonest*. Contrary to the die task, observing individual behavior gives us precise values for each category.

Focusing on significant results, we find that, compared to ticket holders, non–ticket holders in the field display a higher rate of ticket fraud in the game (0.45 versus 0.28; Mann–Whitney test, $p = 0.001$); they are more fully dishonest (Fisher's exact test, $p = 0.033$) and less fully honest ($p = 0.016$). Similarly, compared to self-reported nonfraudsters, self-reported fraudsters have a higher rate of fraud (0.50 versus 0.18, $p < 0.001$); they are more fully ($p = 0.001$) and partially ($p < 0.001$) dishonest, but less fully honest ($p < 0.001$). The fraud rate in the game is significantly larger for subjects who report more frequent frauds in the field (3 or more, 0.64) compared to those who report less frequent frauds (0.37) or no fraud at all ($p < 0.001$ in both cases).[28]

Non–ticket holders who self-report as fraudsters are more dishonest in the game than any other category (fraud rate = 0.55, $p < 0.001$ for each pairwise comparison), followed by ticket holders who self-report as fraudsters (0.44, $p < 0.001$). The remaining two categories, i.e., ticket and non–ticket holders who self-report as nonfraudsters, display similar rates of fraud (0.15 and 0.23, respectively; $p = 0.211$); these rates are smaller than for the other categories ($p < 0.001$). The proportion of fully dishonest subjects is higher among non–ticket holders who self-report as fraudsters compared to non–ticket holders who do not admit they are fraudsters and ticket holders who self-report as fraudsters or as nonfraudsters ($p = 0.009, 0.065$ and 0.001, respectively). Both non-ticket holders and ticket holders who self-report as fraudsters are more partially dishonest and less fully honest than non–ticket holders who self-report as nonfraudsters and ticket holders who self-report as nonfraudsters ($p < 0.05$). Interestingly, ticket and non–ticket holders who self-report as nonfraudsters display similar proportions of fully honest and partially and fully dishonest subjects ($p > 0.1$).

Finally, subjects from the FCO have a higher fraud rate (0.42) than self-reported nonfraudsters ($p < 0.001$), irrespectively of whether the latter hold a validated

Table 5. Fraud Rate and Percentages of Honest and (Partially and Fully) Dishonest Subjects (Public Transportation Game)

	Mean rate of ticket fraud	Fully honest (%)	Partially dishonest (%)	Fully dishonest (%)
All	0.36 (0.37)	38	45	16
Fine Collection Office	0.42 (0.41)	37	40	23
Ticket holders	0.28 (0.34)	45	44	11
Non–ticket holders	0.45 (0.38)	29	49	22
Self-reported nonfraudsters	0.18 (0.31)	68	25	7
Self-reported fraudsters	0.50 (0.35)	14	63	22
1 or 2 frauds/10 trips	0.37 (0.32)	20	68	13
3 or more frauds/10 trips	0.64 (0.32)	8	59	33
No ticket/Self-reported fraudsters	0.55 (0.37)	14	57	29
No ticket/Self-reported nonfraudsters	0.23 (0.33)	61	33	6
Ticket/Self-reported fraudsters	0.44 (0.32)	14	71	15
Ticket/Self-reported nonfraudsters	0.15 (0.29)	71	21	8

Note. Standard deviations are in parentheses.

ticket ($p<0.001$) or not ($p = 0.038$); they are also more fully ($p = 0.025$) and partially dishonest ($p = 0.088$) and less fully honest ($p = 0.001$). Their fraud rate is smaller ($p = 0.082$) and they are more fully honest ($p = 0.012$) than non–ticket holders who admit they are fraudsters.

To test the robustness of these results, we estimate four logit regressions in which the dependent variable is the decision to *not* buy a ticket for a given destination. Since we have multiple observations for each subject, we cluster the standard errors at the individual level.[29] The independent variables are the same as in Table 4, except that we include a dummy for the variable treatment and a control variable for the number of stops needed to reach the destination (from one to seven). Table 6 reports marginal effects and standard errors.

Table 6 reveals that self-reported fraud in the field is a better predictor of dishonesty in this game than not holding a ticket. Compared to self-reported fraudsters, the non–ticket holders who self-report as nonfraudsters behave as honestly as the ticket holders who self-report as nonfraudsters. The coefficients of the two variables relative to self-reported nonfraudsters (holding or not a ticket) differ neither in model (3) nor in model (4) ($p = 0.233$ and $p = 0.562$, respectively). In contrast, the effect of holding versus not holding a ticket reduces the probability of fraud in model (1), but not in model (2) where we control for individual characteristics. These results differ partly from those of the die task. Our interpretation is that the two games do not differ only in terms of context, but also in terms of possible feeling of scrutiny. Indeed, even if anonymity is preserved, individual behavior can be, ex post, verified by the experimenter. This is a crucial difference from the die task, where information about the true outcome of the die rolls remains private. This difference might explain why the subjects who do not hold a ticket but inconsistently self-report as nonfraudsters in the field lie a lot in the die task but not in the public transportation game. Image concerns may explain that these non–ticket holders do not acknowledge they are fraudsters and that they cheat less when they anticipate that we will be able to check behavior ex post.

4.4. Comparison Between Dishonesty in the Die Task and in the Public Transportation Game

As mentioned earlier, the primary objective of our study is not to compare the cheating behavior in the die task and public transportation game, but to test whether the correlation between behavior in the field and in the lab holds in these two different environments. The two

Table 6. Determinants of Fraud in the Public Transportation Game (Logit Regressions)

	(1)		(2)	
	dy/dx	SE	dy/dx	SE
Ticket holder	−0.093**	0.042	−0.063	0.043
Self-reported nonfraudster	−0.292***	0.040	−0.155***	0.050
Fine Collection Office	−0.096	0.069	< 0.001	0.064
Controls for individual characteristics	No		Yes	
Controls for the treatment variables	Yes		Yes	
Number of observations	1,953		1,953	
Pseudo-R^2	0.113		0.220	
df	5		31	
Prob > F	<0.001		<0.001	
	(3)		(4)	
	dy/dx	SE	dy/dx	SE
No ticket and self-reported nonfraudster	−0.286***	0.072	−0.173***	0.066
Ticket and self-reported fraudster	−0.090*	0.048	−0.073	0.053
Ticket and self-reported nonfraudster	−0.386***	0.054	−0.214***	0.066
Fine Collection Office	−0.094	0.071	−0.004	0.065
Controls for individual characteristics	No		Yes	
Controls for the treatment variables	Yes		Yes	
Number of observations	1,953		1,953	
Pseudo-R^2	0.113		0.220	
df	6		32	
Prob > F	<0.001		<0.001	

Notes. Logit regressions with clustered standard errors are shown. The table reports the marginal effects. The "dy/dx" notation indicates the derivative, i.e., the rate of change of the variable y in reaction to a change in variable x. Independent variables include individual characteristics, attitudes toward risk, usage of public transportation, norms, and beliefs. The full regressions are contained in Online Appendix 8. In models (3) and (4), the omitted category is "No ticket and self-reported fraudster."

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

laboratory tasks in fact differ in many respects—most notably in the way dishonesty is measured (at group level in the die task and at individual level in the public transportation game) and whether context is used. Therefore, a direct comparison would be too noisy. That said, it is still useful to provide some tests on how behavior in the public transportation game relates to the one in the die task. In the remainder of this section, we will attempt to statistically quantify the extent to which the lab behavior of our different categories of subjects differs between the public transportation game and the die task.[30]

To make the two measures of lab dishonesty comparable, we construct a standardized measure of dishonesty that is based on the payoff consequences of subjects' decisions.[31] We first compute the expected payoff from fully honest behavior ($\bar{\pi}$), the highest possible payoff ($\pi^{\max}$), and the lowest possible payoff ($\pi^{\min}$).[32] We then calculate the standardized measure of dishonesty (d^s) as follows:

$$d^s = \begin{cases} \dfrac{\pi - \bar{\pi}}{\bar{\pi} - \pi^{\min}} & \pi < \bar{\pi}, \\ \dfrac{\pi - \bar{\pi}}{\pi^{\max} - \bar{\pi}} & \pi \geq \bar{\pi}, \end{cases}$$

where π is the realized payoff.

This standardized measure can take values from −1 (when $\pi = \pi^{\min} < \bar{\pi}$) to +1 (when $\pi = \pi^{\max}$), and is equal to 0 when the realized payoff coincides with the expected payoff from fully honest behavior (i.e., $\pi = \bar{\pi}$). The average standardized measure, computed for a group of subjects, measures how much more money they make compared to a benchmark group of fully honest subjects. In particular, the more d^s is greater than zero, the more dishonest their behavior is.

We can now compare this standardized measure of dishonesty across our categories of subjects and

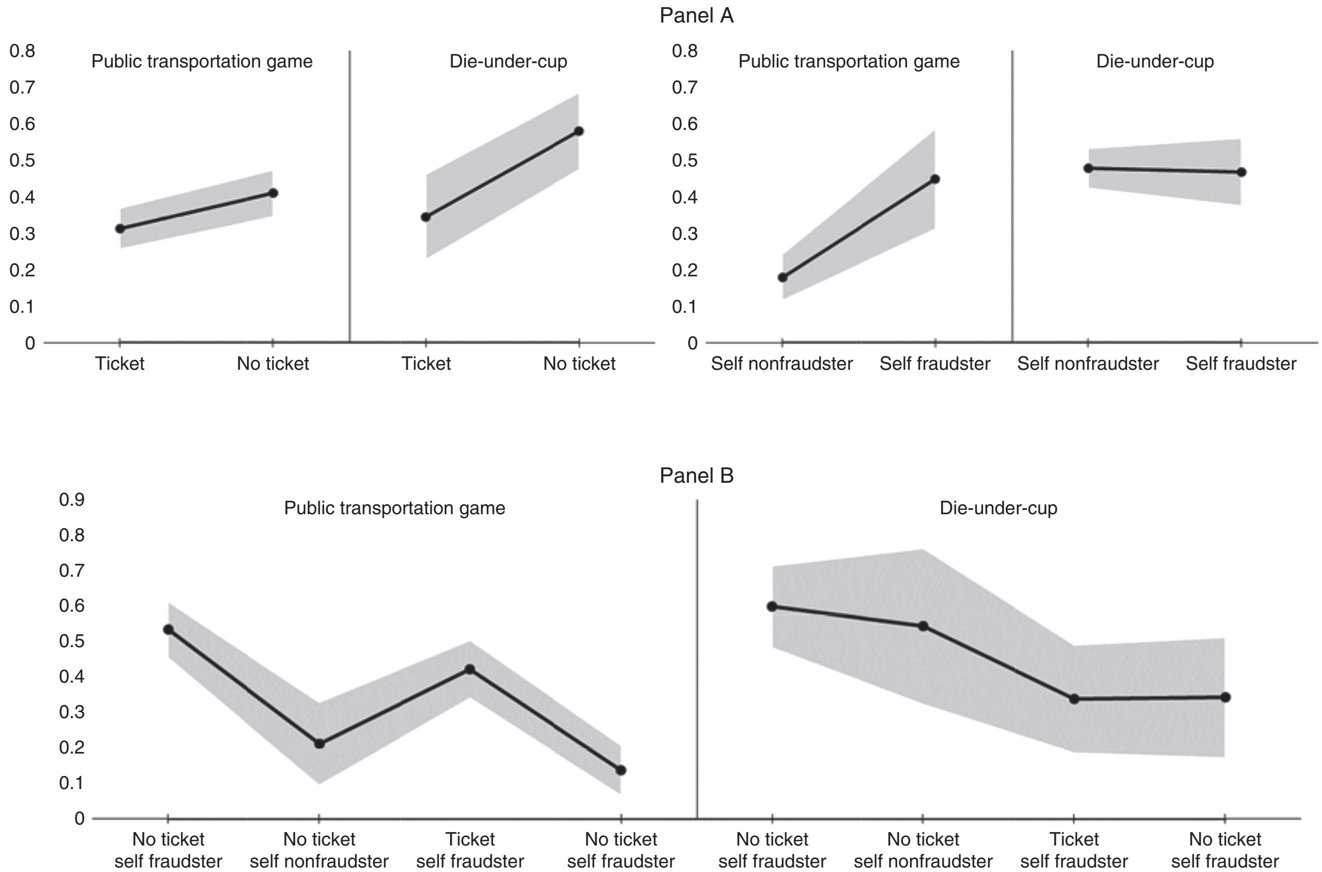
Table 7. Regressions on the Standardized Measure of Dishonesty

	(1)		(2)	
	dy/dx	SE	dy/dx	SE
Ticket holder	−0.235***	0.081	−0.244***	0.088
Self-reported nonfraudster	−0.019	0.085	0.005	0.091
Public transportation game	−0.084	0.070	−0.084	0.072
Fine Collection Office	−0.314**	0.136	−0.249*	0.129
Ticket holder × Public transportation game	0.139	0.094	0.139	0.097
Self-reported nonfraudster × Public transportation game	−0.280***	0.098	−0.280***	0.100
Fine Collection Office × Public transportation game	0.191	0.150	0.191	0.154
Controls for individual characteristics	No		Yes	
Number of observations	558		558	
Pseudo-R^2	0.079		0.144	
df	7		33	
Prob > F	<0.001		<0.001	
	(3)		(4)	
	dy/dx	SE	dy/dx	SE
No ticket and self-reported nonfraudster	−0.055	0.127	−0.061	0.135
Ticket and self-reported fraudster	−0.260***	0.099	−0.291***	0.104
Ticket and self-reported nonfraudster	−0.256**	0.106	−0.239**	0.114
Fine Collection Office	−0.325**	0.138	−0.269**	0.130
Public transportation game	−0.089	0.075	−0.089	0.077
No ticket and self-reported nonfraudster × Public transportation game	−0.266*	0.152	−0.266*	0.156
Ticket and self-reported fraudster × Public transportation game	0.148	0.113	0.148	0.116
Ticket and self-reported nonfraudster × Public transportation game	−0.141	0.121	−0.141	0.124
Fine Collection Office × Public transportation game	0.195	0.152	0.195	0.156
Controls for individual characteristics	No		Yes	
Number of observations	558		558	
Pseudo-R^2	0.08		0.146	
df	9		35	
Prob > F	<0.001		<0.001	

Notes. OLS regressions with clustered standard errors are shown. The "dy/dx" notation indicates the derivative, i.e., the rate of change of the variable y in reaction to a change in variable x. Independent variables include individual characteristics, attitudes toward risk, usage of public transportation, norms, and beliefs. The full regressions are available from the authors upon request. In models (3) and (4), the omitted category is "No ticket and self-reported fraudster."

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Figure 2. Average Marginal Effects of the Standardized Measure of Dishonesty



Note. The shaded area identifies the 95% confidence intervals.

between our two laboratory tasks by means of a regression analysis. We employ OLS regressions with standard errors clustered on each subject to control for the dependence of the decisions between the die task and the public transportation game. Explanatory variables include all those used in previous regressions, a dummy variable for the task (die task = 0; public transportation game = 1), and interactions between this dummy and the field measures of dishonesty. Table 7 reports the results of these regressions. Figure 2 displays the average marginal effects for our different categories of subjects distinguishing between behavior in the die task and in the public transportation game. More specifically, panel A reports the average marginal effects for ticket holders, non–ticket holders, self-reported fraudsters, and self-reported nonfraudsters, whereas panel B displays the average marginal effects for self-reported fraudsters who do not hold a validated ticket, self-reported fraudsters who hold a validated ticket, self-reported nonfraudsters who do not hold a validated ticket, and self-reported nonfraudsters who hold a validated ticket.

The results support our previous findings. In particular, while not holding a ticket is a good predictor of lab behavior, especially in the die task but also in the public transportation game, self-reported fraud in the field is a good predictor of lab dishonesty only in the public transportation game.[33] In addition, we confirm that in the public transportation game, non–ticket holders who self-report as nonfraudsters behave as honestly as the ticket holders who self-report as nonfraudsters. All this supports our interpretation that scrutiny matters for image-concerned subjects (i.e., those who did not hold a valid ticket and are uncomfortable to admit they are fraudsters).

5. A Complementary Study with Students

So far, we have provided evidence of the external validity of laboratory tasks performed by a pool of nonstudent subjects. Most lab experiments are, however, conducted with students. It is thus legitimate to ask whether our results hold with undergraduate students in a typical lab environment. To bridge our artefactual field experiment with a conventional lab experiment, we invited 192 university students to perform the same die-under-cup task in the context of a typical lab experiment.[34] The procedures were similar to the field experiment, except for a few changes. First, subjects were recruited online using the H-root software (Bock et al. 2014). Second, the experiment was conducted at GATE-LAB, Lyon. Last, we collected a different measure of field dishonesty, which consists in overpaying subjects at the end of the session and seeing whether they report or not the "mistake" in the payment. This manipulation has been used by Gneezy et al. (2014) and Potters and Stoop (2016) in the lab and by Azar et al. (2013) in the field.[35] Subjects were paid their earnings and extra €5 at their computer station using envelopes. The earnings of the experiment (excluding the overpayment) remained displayed on the computer screen of each subject during the payment procedures. Subjects were also reminded to check the money they received before leaving the laboratory. Those who reported the mistake in the payment were classified as honest and the others as dishonest.[36]

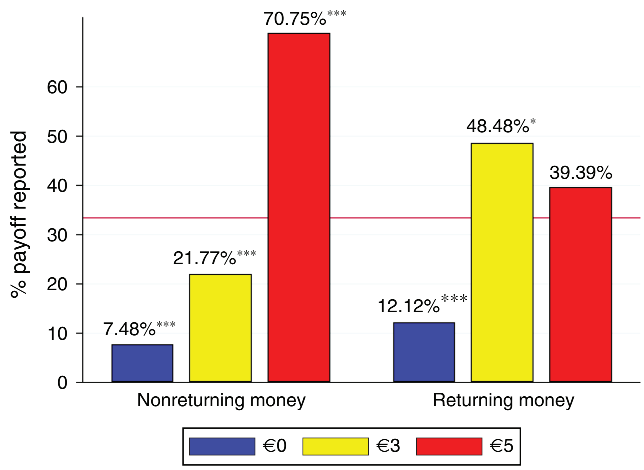
Result 6 summarizes our findings:

Result 6. A conventional lab experiment with undergraduate students replicates (a) the diversity of cheating behavior in the die task observed with passengers of public transportation and (b) the relation between the lab task and a field measure of dishonesty.

Support for Result 6. The distribution of the die outcomes reported by the students is not uniform (χ^2 test, $p < 0.001$); indeed, the percentages of reds, yellows, and blues are 65.10%, 26.56%, and 8.33%, respectively, showing evidence of widespread cheating. This distribution is also statistically different from the one observed in our field experiment (Fisher's exact test, $p = 0.018$). In particular, students report more reds ($p = 0.06$) and fewer yellows ($p = 0.021$) than people recruited in the field (but not blues, $p = 0.352$).

Next, we test whether cheating in the die task is associated with our measure of dishonesty in the field. Figure 3 displays the distributions of the die outcomes reported by the students who do not return the extra earnings, on the left, and by those who return them, on the right. Of the 192 subjects who took part in the experiment, 12 (6.25%) were not treated, 147 (76.56%) did not return the extra earnings, and 33 (17.19%) returned them. The distribution reported by those who do not return the overpayment differs significantly from a uniform distribution (χ^2 test, $p<0.001$). These students report more reds, fewer yellows, and fewer blues than the expected values of 33% (Fisher's exact test, $p < 0.001$ for reds and blues, and $p = 0.003$ for yellows). Turning to those who return the overpayment, the distribution of their reported outcomes also differs from the uniform distribution (χ^2 test, $p = 0.029$). In particular, they report fewer blues and more yellows than the expected values of 33% (Fisher's exact test, $p = 0.009$ for blues, and $p = 0.094$ for yellows; for reds,

Figure 3. (Color online) Reported Outcomes of the First Die Roll and Return of Money by Students



Note. The horizontal line identifies the uniform distribution and is located at 33.33%.

$^{*}p < 0.1$; $^{***}p < 0.01$ (two-sided binomial tests that the observed percentage differs from 33.33%).

the test gives insignificant results, $p = 0.464$). However, their distribution is statistically different from the one reported by those who do not return the extra earnings ($p = 0.002$). In particular, the students who return the overpayment report significantly fewer reds (39.39% versus 70.75%) and more yellows (48.48% versus 21.77%) than those students who did not return the extra money ($p = 0.001$ and 0.004, respectively). They also report more blues (12.12% versus 7.48%), but the difference does not achieve statistical significance ($p = 0.482$).

These results are similar to those obtained with the passengers of public transportation and they confirm that simple laboratory tasks measuring dishonesty are able to capture important aspects of moral flexibility in the field.

6. Discussion and Conclusion

The data from our artefactual field experiment with a diversified sample of passengers of a public transport service confirm that, when given the opportunity to increase their earnings by behaving dishonestly in laboratory tasks without any risk of detection, some individuals follow the moral principle of honesty, while others act partially dishonestly and others cheat fully to maximize their earnings. This diversity of moral behavior is observed regardless of whether the laboratory task is abstract (die task) or contextualized (public transportation game). It confirms previous studies (Fischbacher and Föllmi-Heusi 2013, Gibson et al. 2013). The novelty of our study is to provide evidence that this diversity of behavior in the laboratory tasks replicates the diversity of the degrees of moral firmness of the same individuals in their day-to-day life, for incentives to fraud of comparable monetary value.

Indeed, at the aggregate level, the group of passengers who cannot present a validated ticket and who self-report as frequent fare dodgers tends to act fully dishonestly in the die task. In contrast, the group of passengers who hold a validated ticket and self-report as never evading fares reports honestly the die outcomes more frequently than the other groups. At the individual level, we show that fare dodgers in real life behave also less honestly in the transportation game, while people who self-report as non–fare dodgers are less likely to evade. These major findings indicate that simple lab experiments can deliver externally valid results on rule violation by ordinary populations. They suggest that norm violators have lower intrinsic ethical values than others (since when they cheat they are more likely to do it in our different contexts). However, we are cautious not to conclude that context does not matter in general. In particular, in our study we consider cheating for a given level of incentives that is relatively similar in the field and in the lab. An interesting extension would be to measure the robustness of our findings to a manipulation of the level of incentives. The meta-analysis of Abeler et al. (2016) on studies based on reports of a random draw (dice or coins) concludes that reporting behavior is not affected by the manipulation of the incentive to lie. On the other hand, Kajackaite and Gneezy (2015) show that the insensitivity of lying to incentives results from the subjects' concern about being exposed as cheaters. In contrast, the impact of variations of incentives on cheating behavior in real settings is not well understood yet. Exploring the influence of parallel manipulations of the incentives to lie in the lab and in the field is beyond the scope of this study but would be extremely useful. It would be also interesting to test whether the correlation that we observe still holds when individuals have benefited from positive income shocks in their life.

We chose a public transportation environment because of its economic relevance (i.e., fare dodging concerns ordinary people and entails dramatic revenue losses for transport companies) and of the possibility to obtain a relatively simple and reliable measure of dishonesty in the field. One can naturally wonder whether our results could generalize to other environments. This is why we conducted a complementary experiment with students. In contrast to studies on cooperation in social dilemmas conducted with both fishermen and students in the lab and in the field (Stoop et al. 2012), but consistent with most studies on other-regarding preferences, preferences for efficiency, peer effects, or market behavior (see surveys by Fréchette 2015 and Herbst and Mas 2015), we find that the students' dishonest behavior largely replicates that of ordinary people. Indeed, our second study confirms the within-subject correlation between dishonesty in the laboratory and in a real setting. Future studies could test the extent to which our findings might extend to other domains of dishonesty.

In terms of methodological implications, by showing that cheating in the lab and cheating in the field—as measured by objective measures of dishonesty and self-reports—are correlated for ordinary people as well as for students, our findings provide support for the external validity of lab experiments in a domain that is not easily observable in the field because of its secretive nature. There is, however, an important caveat. As image-concerned people are reluctant to admit they are cheaters, in the lab, the same persons may inflate their degree of honesty if they feel scrutinized. This is similar to what may happen in experiments studying prosocial behavior (e.g., List 2006, Alpizar et al. 2008, Benz and Meier 2008). Our results emphasize the importance of preserving the confidentiality of decisions, especially when studying honesty. Indeed, we found that self-reports have a higher predictive value in the transportation game than in the die task. This may be driven by the fact that in the former the experimenter is able, ex post, to detect dishonesty at the individual level. Subjects who self-reported as never cheating although they were not able to show a valid ticket behaved like self-reported fare dodgers in the die task (where there is no scrutiny), but like ticket holders in the transportation game. Using tasks where decisions can be verified may underestimate the dishonesty of image-concerned individuals who conceal their preferences when dishonesty can be identified even with no risk of sanction.

We have also shown that individuals who have just been fined in a real setting behave more honestly in the die task than other fare dodgers, which goes against the notion that people try to recover their losses. This could be driven by self-selection (those who come to the FCO are more "honest" than other fare dodgers who never pay their fines), by the reactivation of the moral norm of honesty, or by the fact that having just paid a fine emphasized the cost of cheating. Our design and the small size of the sample do not allow us to disentangle precisely between these various explanations. Moreover, we ignore the reason of their sanction and the level of their fine. However, the fact that these subjects cheat as often as other fare dodgers in the transportation game does not support the first explanation. It also suggests that the reactivation of the moral norm, if it exists, does not resist the priming of the identity of fare dodgers by the introduction of context in the transportation game. While the role of identity has already been shown (Cohn et al. 2014, 2015), an extension could usefully investigate the dynamics of self-image and identity in the domain of dishonesty.

If the lab results can identify reliably both directional and quantitative effects regarding rule violation in the field, this will be very useful to inform on the underlying mechanisms of dishonesty, but will certainly be even more important in predicting how people will react to policy interventions. As developed by Camerer (2015), the generalizability of lab results is crucial if one adopts a policy view. The parallelism that we have identified in this study should increase our confidence in the use of laboratory methods to investigate which manipulation of incentives, sanctions, and moral appeals in the lab is better able to deter dishonest behavior in the field.

Acknowledgments

The authors are extremely grateful to P. Jacquesson, R. Sauter, and F. German from Keolis and to D. Aubaret and L. Clerc-Gagnoux from Kisio for enabling the realization of this experiment with passengers of the public transportation company in Lyon. For very useful comments, the authors thank three reviewers, A. Cohn, E. Fehr, S. Shalvi, and R. Weber, as well as seminar participants at the University of Chicago, the University of Erlangen-Nuremberg, the conference of the Association for the Development of Research in Economics and Statistics in Paris, the Meeting of the International Meeting on Experimental and Behavioral Social Sciences in Toulouse, the Economic Science Association (ESA) World Meeting in Sydney, the ESA European meeting in Heidelberg, the 6th Conference of the French Association of Experimental Economics (ASFEE) in Paris, the 12th Workshop on Social Economy for Young Economists at the University of Bologna, the 5th Zurich Workshop in Economics in Lucerne, the GATE-LAB Workshop on Social Norms and Moral Norms in Lyon, and the Behavioral and Experimental Economics Lyon-Toulouse workshop. They thank Q. Thevenet for programming the experiment and S. Belhadj, L. Charroin, and A. Solda for valuable research assistance.

Endnotes

[1] Irlenbusch and Villeval (2015) review almost 60 articles related to dishonesty published in economic journals between 2012 and early 2015. The meta-analysis of Abeler et al. (2016) includes 58 studies on lying based on the reports of random draws (coins or dice). Of course, there is an older experimental literature on tax evasion (for a review, see Torgler 2007) and cheap talk games with strategic information transmission (e.g., Crawford and Sobel 1982, Gneezy 2005, Sutter 2009) or promises about future behavior (e.g., Charness and Dufwenberg 2006).

[2] Subjects frequently do not deceive or lie in a way that maximize their earnings (Gneezy 2005, Shalvi et al. 2011, Fischbacher and Föllmi-Heusi 2013). Possible explanations are that people want to preserve a good self- or social image (Mazar et al. 2008, Gino et al. 2009, Shalvi et al. 2011, Fischbacher and Föllmi-Heusi 2013) or they follow moral norms (Vanberg 2008, Lundquist et al. 2009, Pruckner and Sausgruber 2013). Subjects also condition their decision to lie on the effect that their behavior will have on others (Brandts and Charness 2003, Charness and Dufwenberg 2006, Ellingsen et al. 2010, Maggian and Villeval 2016), suggesting a role for social preferences and guilt. Dishonesty is also sensitive to occupational norms (Cohn et al. 2014) and incentive schemes (e.g., Conrads et al. 2013, 2014; Charness et al. 2014). Gächter and Schultz (2016) demonstrated that individual dishonesty in the lab is also directly influenced by the prevalence of rule violations in the society.

[3] Fare dodging and fare evasion are used interchangeably throughout this paper. They are considered as a crime and are defined as using public transports without purchasing the required ticket or without validating one's ticket at the machine.

[4] Surprisingly, economic studies on fare evasion are very few (Boyd et al. 1989, Kooreman 1993, Nikiforakis 2007, Bucciol et al. 2013, Buehler et al. 2014). More can be found in criminology (e.g., van Andel 1989, Hauber et al. 1996, Smith and Clarke 2000, Killias et al. 2009). Data from the field are usually not made available by companies. Fragmented information comes from newspapers but data are difficult to compare. A report of the French Auditing Court (Cour des Comptes 2016) on transportation in the region Paris-Ile de France indicates that the fraud rate on buses and tramways has increased from about 9% in 2009 to about 14% in 2014, but acknowledges that these percentages are underestimated. It indicates that in 2013, the number of trips without a valid ticket amounted to 14 million trips on the RER (Réseau Express Régional), 23 million on tramways, 84 million on the metro, and 123 million on buses.

[5] Economists are typically reluctant to use self-reports. In our case, this measure is used as a complement and can be correlated with our objective measure of fraud. Moreover, studies in psychology show a correlation between self-reports on unethical tendencies and actual lying (Halevy et al. 2014, Zimerman et al. 2014).

[6] People may not reveal their true preferences if their behavior can be observed. To circumvent this problem, economists and psychologists developed recently simple techniques to measure dishonesty in total privacy. Besides reporting the outcome of a dice roll in private (e.g., Shalvi et al. 2011, Fischbacher and Föllmi-Heusi 2013, Hao and Houser 2013, Gächter and Schultz 2016), they include coin flips (e.g., Bucciol and Piovesan 2011, Houser et al. 2012, Abeler et al. 2014) and reports on real-effort outcomes (e.g., Mazar et al. 2008, Rosaz and Villeval 2012). In mind games, the experimenter knows the result of the random draw, but payment depends on the self-reported accuracy of a prior prediction (Jiang 2013, Kajackaite and Gneezy 2015).

[7] By diversifying their subject pool, some studies on dishonesty found substantial differences between student and nonstudent subjects (Utikal and Fischbacher 2013, Cohn et al. 2014), while others found no difference (Abeler et al. 2014). This suggests that more investigation is needed.

[8] This is an old debated topic in psychology. Hartshorne and May (1928) claim that cheating in one context cannot predict cheating in other contexts, while more recent studies defend the generality of moral behavior (Burton 1963).

[9] Indeed, priming methods may have some limitations. For example, the behavior change observed under priming may not occur actually in the real world. This is the case if the priming manipulation induces an experimenter demand effect (Zizzo 2012). Moreover, the effect induced by priming does not tell whether the task is, in the first place, able to capture differences in dishonesty in the field.

[10] Using a similar approach, Cohn and Maréchal (2015) show that cheating in a coin-tossing task predicts school misbehavior of middle and high school students. A major difference with our study is that the violation of behavioral rules differs from dishonesty. Other differences include the fact that we compare different tasks.

[11] They also show that students with above median dice score are more likely to be willing to work in the public sector where corruption is high. Barfort et al. (2015) find the opposite in Denmark. However, the latter do not observe dishonesty in the field.

[12] Hanna and Wang (2015) look at a very specific population, composed mainly of women (95%), and pay subjects with candies. The fact that some nurses have been exposed to a biometric monitoring with the goal of reducing absenteeism may also have affected the measurement of dishonesty in the experimental task.

[13] Recruiters were also instructed not to invite both members of the same couple or several friends simultaneously. They were told not

to recruit people who spontaneously offered themselves to take part in the study to avoid potential contamination from subjects having already participated.

14 For the third part, in this paper we consider only the first period. The fourth part consisted of a dynamic public transportation game in which subjects had to decide whether to buy a ticket or not and in which we manipulated the frequency and regularity of audits and the information about the future occurrence of audits. The results of this fourth part are reported in Dai et al. (2017).

15 Subjects have to report whether they are fully prepared to take risks or whether they prefer to avoid risks on a scale between 0 ("not at all willing to take risks") and 10 ("very willing to take risks"), using the Dohmen et al. (2011) procedure. The second question is similar but in the context of financial decisions.

16 The second roll is presented as allowing the subjects to check that the die is fair.

17 After playing period 1, subjects discovered four subsequent periods in this game. In periods 2 to 5, which we do not consider in this paper, we introduced various common knowledge probabilities of being controlled and fined. The results of the other periods are available upon request.

18 It is unlikely that subjects trashed their ticket in the bus/tram before being approached by the recruiters since this behavior is extremely unusual in Lyon. One can argue that some subjects might have thrown out their ticket on the way to the lab. Although we cannot entirely exclude this possibility, we believe it is very unlikely. Subjects had to walk only 60 meters from the stop, and had been just advised not to trash the ticket. None of the recruiters reported such behavior. Finally, even if few subjects did trash their ticket and were wrongly classified as dishonest, it would only make it more difficult to detect differences in behavior between ticket holders and non–ticket holders, and, therefore, it would strengthen our results.

19 As mentioned earlier, this only reduces the distinction between ticket holders and non–ticket holders, making any difference we detect even stronger.

20 We believe that allowing people to provide a nonbinary answer to this question reduces the number of subjects who self-report as nonfraudsters when they are actually fraudsters.

21 We cannot provide these statistics for the FCO subjects since they may have not used public transport.

22 The high fraud rate is also due to the recruitment of passengers without a monthly pass. Recruiting metro users would have also led to a much lower fraud rate because the access to metro is protected by barriers. The rate of self-reported fraud is consistent with a 2011 survey conducted by OpinionWay in Lyon, where it was found that 55% of respondents reported to sometimes travel without a valid ticket (Keolis 2014). These statistics can also be compared to the percentage of non–ticket holders (42.88%) found by Bucciol et al. (2013) in a medium-sized city in Italy. Italy and France display among the highest rates of fare evasion in Europe (Bonfanti and Wagenknecht 2010).

23 Throughout this paper, statistical tests are two-sided unless specified otherwise. Each subject gives one independent observation. To compare the distributions of the reported outcomes with the theoretical distributions, we use χ^2 goodness-of-fit tests. Kolmogorov–Smirnov tests (available upon request) give similar results, but they are less desirable since they assume continuous distributions.

24 Self-reported fraudsters without a ticket report significantly higher outcomes than self-reported fraudsters and nonfraudsters with a ticket ($p = 0.039$ and 0.018, respectively). Self-reported nonfraudsters without a ticket report higher outcomes than self-reported fraudsters ($p = 0.068$).

25 Being a male, having an income lower than €1,000, having a high school education or lower, and living in the city center or in a lower class neighborhood all increase the probability to be a fare dodger (non–ticket holders and/or self-reported fraudsters). Actual fare dodgers have also weaker moral norms of honesty (higher approval, more flexibility in the definition of dishonesty, higher beliefs about the fraud rate in public transportation in Lyon; see Online Appendix 5).

26 Ordered logit regressions give qualitatively similar results. They are reported in Online Appendix 10.

27 In Online Appendix 7, we complement this study with an analysis of self-justification in the die task. In other words, we investigate whether subjects use the information on the second die roll as a self-justification to lie by reporting the best outcome of the two die rolls (see Shalvi et al. 2011).

28 These differences across subject categories hold when analyzing the other periods of this game when we introduce positive probabilities of detection. Results are available upon request.

29 Random-effects logit models give qualitatively similar results (available upon request).

30 In Online Appendix 9, we report additional analysis on the correlation between behavior in the public transportation game and in the die task.

31 This analysis builds upon some of the original techniques used in a recent meta study by Abeler et al. (2016).

32 In the die task, fully honest behavior is equivalent to truthfully reporting the outcome of the die. Hence, the expected payoff from fully honest behavior is equal to $(0+3+5)/3 = €2.6\bar{7}$. The highest (lowest) possible payoff is equal to €5 (€0). In the public transportation game, since subjects make repeated decisions, we consider average payoffs. Fully honest behavior means always buying the ticket. The expected payoff from this behavior is equal to $8 - 1.7 = €6.3$, where 8 is the average benefit from using the fictional bus and 1.7 the cost of the ticket. The highest (lowest) possible payoff is equal to €8 (€6.3).

33 In previous regressions (see Table 4), we find that in the die task, self-reported fraudsters tend to overreport the middle outcome compared to self-reported nonfraudsters, suggesting that they are slightly more dishonest than self-reported nonfraudsters. We do not find this in Table 7.

34 The task was added at the beginning of another unrelated experiment. As in our artefactual field experiment, we collected data on the subjects' risk preferences before playing the die-under-cup task. Students were recruited from the local engineering and business schools.

35 In fact, 12 of our 192 subjects were not treated (we did not pay them the extra €5). Indeed, in the first two sessions (out of 13) we treated only half of the subjects to make sure that everything worked smoothly.

36 We acknowledge that this is a crude measure of dishonesty. First, lazy or naïve subjects could leave the lab without checking their money. Second, we measure a mild form of dishonesty that requires passive cheating (in contrast to the die task where cheating requires active misreporting). All this would, however, only make it more difficult to detect a correlation between field and lab dishonesty. Hence, if we find one, as we do, this result is even stronger. Moreover, passive cheating is relatively similar to fare dodging where individuals do not buy a ticket if they want to cheat, and the monetary stake of the fraud is in the same range of values.

References

Abeler J, Becker A, Falk A (2014) Representative evidence on lying costs. *J. Public Econom.* 113(May):96–104.

Abeler J, Nosenzo D, Raymond D (2016) Preferences for truth-telling—A meta-study. IZA Discussion Paper 10188, Bonn.

Alm J, Bloomquist KM, McKee M (2015) On the external validity of laboratory tax compliance experiments. *Econom. Inquiry* 53(2):1170–1186.

Alpizar F, Carlsson F, Johansson-Stenman O (2008) Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *J. Public Econom.* 92:1047–1060.

Armantier O, Boly A (2013) Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. *Econom. J.* 123(573):1168–1187.

Azar OH, Shira Y, Bar-Eli M (2013) Do customers return excessive change in a restaurant? *J. Econom. Behav. Organ.* 93(September):219–226.

Barfort S, Harmon N, Hjorth F, Olsen AL (2015) Dishonesty and selection into public service in Denmark: Who runs the world's lest corrupt public sector? Working paper, University of Copenhagen, Copenhagen.

Becker GS (1968) Crime and punishment: An economic approach. *J. Political Econom.* 76(2):169–217.

Benz M, Meier S (2008) Do people behave in experiments as in the field? Evidence from donations. *Experiment. Econom.* 11(3): 268–281.

Bock O, Baetge I, Nicklisch A (2014) hroot: Hamburg Registration and Organization Online Tool. *Eur. Econom. Rev.* 71(October):117–120.

Bonfanti G, Wagenknecht T (2010) Human factors reduce aggression and fare evasion. *Public Transport Internat.* 59(1):28–32.

Boyd C, Martini C, Rickard J, Russell A (1989) Fare evasion and noncompliance. *J. Transport Econom. Policy* 23(2):189–197.

Brandts J, Charness G (2003) Truth or consequence: An experiment. *Management Sci.* 49(1):116–130.

Bucciol A, Piovesan M (2011) Luck or cheating? A field experiment on honesty with children. *J. Econom. Psych.* 32(1):73–78.

Bucciol A, Landini F, Piovesan M (2013) Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *J. Econom. Behav. Organ.* 93:248–257.

Buehler S, Halbheer D, Lechner M (2014) Payment evasion. Working paper, University of St. Gallen, St. Gallen, Switzerland.

Burton RV (1963) Generality of honesty reconsidered. *Psych. Rev.* 70(6):481–499.

Camerer CF (2015) The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Frechette GR, Schotter A, eds. *Handbook of Experimental Economic Methodology* (Oxford University Press, Oxford, UK), 249–295.

Cardenas JC, Carpenter J (2013) Risk attitudes and economic well-being in Latin America. *J. Development Econom.* 103(July):52–61.

Charness G, Dufwenberg M (2006) Promises and partnership. *Econometrica* 74(6):1579–1601.

Charness G, Fehr E (2015) From the lab to the real world. *Science* 350(6260):512–513.

Charness G, Masclet D, Villeval MC (2014) The dark side of competition for status. *Management Sci.* 60(1):38–55.

Choo L, Fonseca MA, Myles GD (2014) Do students behave like real taxpayers? Experimental evidence on taxpayer compliance from the lab and from the field. Working paper, University of Exeter, Exeter, UK.

Cohn A, Maréchal MA (2015) Laboratory measure of cheating predict misbehavior at school. Working Paper 205, University of Zurich, Zurich.

Cohn A, Fehr E, Maréchal MA (2014) Business culture and dishonesty in the banking industry. *Nature* 516:86–89.

Cohn A, Maréchal MA, Noll T (2015) Bad boys: How criminal identity salience affects rule violation. *Rev. Econom. Stud.* 82(4): 1289–1308.

Conrads J, Irlenbusch B, Rilke RM, Walkowitz G (2013) Lying and team incentives. *J. Econom. Psych.* 34(1):1–7.

Conrads J, Irlenbusch B, Rilke RM, Schielke A, Walkowitz G (2014) Honesty in tournaments. *Econom. Lett.* 123(1):90–93.

Cooper DJ, Kagel JH, Lo W, Gu QL (1999) Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *Amer. Econom. Rev.* 89(4): 781–804.

Cour des Comptes (2016) La lutte contre la fraude dans les transports urbains en Île-de-France: Un échec collectif. Rapport public annuel 2016, Cour des Comptes, Paris, 537–577.

Crawford V, Sobel J (1982) Strategic information transmission. *Econometrica* 50(6):1431–1451.

Dai Z, Galeotti F, Villeval MC (2017) The efficiency of crackdowns: A lab-in-the-field experiment in public transportations. *Theory Decision* 82(2):249–271.

Dai Z, Hogarth RM, Villeval MC (2015) Ambiguity on audits and cooperation in a public goods game. *Eur. Econom. Rev.* 74(February): 146–162.

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econom. Assoc.* 9(3): 522–550.

Donohue P (2010) Millions of fare-beaters take MTA for an $8 million bus ride. *Daily News* (March 17), http://www.nydailynews.com/news/crime/millions-fare-beaters-mta-8-million-bus-ride-article-1.175537.

Eckel CC, Grossman PJ (2008) Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *J. Econom. Behav. Organ.* 68(1):1–17.

Eckel CC, Grossman PJ, Johnson CA, de Oliveira AC, Rojas C, Wilson RK (2012) School environment and risk preferences: Experimental evidence. *J. Risk Uncertainty* 45(3):265–292.

Ellingsen T, Johannesson M, Tjøtta S, Torsvik G (2010) Testing guilt aversion. *Games Econom. Behav.* 68(1):95–107.

Erat S, Gneezy U (2012) White lies. *Management Sci.* 58(4):723–733.

Falk A, Heckman J (2009) Lab experiments are a major source of knowledge in the social sciences. *Science* 326(5952):535–538.

Fischbacher U, Föllmi-Heusi F (2013) Lies in disguise: An experimental study on cheating. *J. Eur. Econom. Assoc.* 11(3):525–547.

Fréchette GR (2015) Laboratory experiments: Professionals versus students. Fréchette GR, Schotter A, eds. *Handbook of Experimental Economic Methodology* (Oxford University Press, Oxford, UK), 360–390.

Fréchette GR, Schotter A, eds. (2015) *Handbook of Experimental Economic Methodology* (Oxford University Press, Oxford, UK).

Fürst E (2015) Free riders and ticket fraud in public transport: A Delphi analysis. Conference paper, Association for European Transport, Henley-in-Arden, UK. http://abstracts.aetransport.org/paper/index/id/3905/confid/18.

Gächter S, Schultz JF (2016) Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531(7595): 496–499.

Gibson R, Tanner C, Wagner AF (2013) Preferences for truthfulness: Heterogeneity among and within individuals. *Amer. Econom. Rev.* 103(1):532–548.

Gino F, Ayal S, Ariely D (2009) Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psych. Sci.* 20(3):393–398.

Gneezy U (2005) Deception: The role of consequences. *Amer. Econom. Rev.* 95:384–394.

Gneezy U, Imas A, Madarasz K (2014) Conscience accounting: Emotion dynamics and social behavior. *Management Science* 60(11):2645–2658.

Halevy R, Shalvi S, Verschuere B (2014) Being honest about dishonesty: Correlating self-reports and actual lying. *Human Comm. Res.* 40(1):54–72.

Hanna R, Wang SY (2015) Dishonesty and selection into public service. Working paper, Harvard University, Boston.

Hao L, Houser D (2013) Perceptions, intentions, and cheating. Working paper, Interdisciplinary Center for Economic Science, George Mason University, Fairfax, VA.

Harrison GW, List JA (2004) Field experiments. *J. Econom. Literature* 42(4):1009–1055.

Hartshorne H, May MA (1928) *Studies in the Nature of Character:, Studies in Deceit*, Vol. 1 (Macmillan, New York).

Hauber A, Hofstra B, Toornvliet L, Zandbergen A (1996) Some new forms of functional social control in the Netherlands and their effects. *British J. Criminology* 36(2):199–219.

Herbst D, Mas A (2015) Peer effects on worker output in the laboratory generalize to the field. *Science* 350(6260):545–549.

Houser D, Vetter S, Winter J (2012) Fairness and cheating. *Eur. Econom. Rev.* 56(8):1645–1655.

Irlenbusch B, Villeval MC (2015) Behavioral ethics: How psychology influenced economics and how economics might inform psychology? *Current Opinion Psych.* 6:87–92.

IRS (2016) Tax gap estimates for tax years 2008–2010. Report, Internal Revenue Service, Washington, DC. https://www.irs.gov/PUP/newsroom/tax%20gap%20estimates%20for%202008%20through%202010.pdf.

Jiang T (2013) Cheating in mind games: The subtlety of rules matters. *J. Econom. Behav. Organ.* 93:328–336.

Kajackaite A, Gneezy U (2015) Lying costs and incentives. Discussion Paper, University of California, San Diego, La Jolla.

Kastlunger B, Kirchler E, Mittone L, Pitters J (2009) Sequences of audits, tax compliance, and tax paying strategies. *J. Econom. Psych.* 30:405–418.

Keolis (2014) Fraude: Comment lutter? Accessed December 28, 2015, http://www.keolis.com/.

Killias M, Scheidegger D, Nordenson P (2009) The effects of increasing the certainty of punishment: A field experiment on public transportation. *Eur. J. Criminology* 6(5):387–400.

Kooreman P (1993) Fare evasion as a result of expected utility maximization. *J. Transport Econom. Policy* 27(1):69–74.

Levitt S, List JA (2007) What do laboratory experiments social preferences reveal about the real world. *J. Econom. Perspect.* 21(2): 153–174.

List JA (2006) The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *J. Political Econom.* 114(1):1–37.

List JA (2009) The economics of open air markets. NBER Working Paper 15420, National Bureau of Economic Research, Cambridge, MA.

Lundquist T, Ellingsen T, Gribbe E, Johannesson M (2009) The aversion to lying. *J. Econom. Behav. Organ.* 70:81–92.

Maggian V, Villeval MC (2016) Social preferences and lying aversion in children. *Experiment. Econom.* 19(3):663–685.

Mazar N, Ariely D (2006) Dishonesty in everyday life and its policy implications. *J. Public Policy Marketing* 25(1):117–126.

Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: A theory of self-concept maintenance. *J. Marketing Res.* 45(6): 633–644.

Nikiforakis N (2007) Fines for young public transport offenders: Insights from the economics literature. Unpublished paper, University of Melbourne, Parkville, Australia.

Potters J, Stoop J (2016) Do cheaters in the lab also cheat in the field? *Eur. Econom. Rev.* 87(August):26–33.

Pruckner GJ, Sausgruber R (2013) Honesty on the streets: A field study on newspaper purchasing. *J. Eur. Econom. Assoc.* 11(3): 661–679.

Rosaz J, Villeval MC (2012) Lies and biased evaluation: A real-effort experiment. *J. Econom. Behav. Organ.* 84(2):537–549.

Shalvi S, Dana J, Handgraaf MJJ, De Dreu CKW (2011) Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Human Decision Processes* 115:181–190.

Smith MJ, Clarke RV (2000) Crime and public transport. *Crime Justice: Rev. Res.* 27:169–233.

Stoop J, Noussair CN, Van Soest D (2012) From the lab to the field: Cooperation among fishermen. *J. Political Econom.* 120(6): 1027–1056.

Sutter M (2009) Deception through telling the truth?! Experimental evidence from individuals and teams. *Econom. J.* 119:47–60.

Torgler B (2007) *Tax Compliance and Tax Morale. A Theoretical and Empirical Analysis* (Edward Elgar, Cheltenham, UK).

Utikal V, Fischbacher U (2013) Disadvantageous lies in individual decisions. *J. Econom. Behav. Organ.* 85:108–111.

van Andel H (1989) Crime Prevention that works: The care of public transport in the Netherlands. *British J. Criminology* 29(1): 47–56.

Vanberg C (2008) Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76(6):1467–1480.

Zimerman L, Shalvi S, Bereby-Meyer Y (2014) Self-reported ethical risk taking tendencies predict actual dishonesty. *Judgment and Decision Making* 9(1):58–64.

Zizzo DJ (2012) Inducing natural group identity: A RDP analysis. CBESS Discussion Paper 12-01, Centre for Behavioural and Experimental Social Science, University of East Anglia, Norwich, UK.