Fraud Deterrence Institutions Reduce Intrinsic Honesty

Fabio Galeotti, Valeria Maggian, Marie Claire Villeval

GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 1924 – September 2019

Fraud Deterrence Institutions Reduce Intrinsic Honesty

Fabio Galeotti, Valeria Maggian, Marie Claire Villeval

Abstract:

Deterrence institutions are widely used in modern societies to discourage rule violations but whether they have an impact beyond their immediate scope of application is usually ignored. Using a natural field experiment, we show that they affect intrinsic honesty across contexts. We identified fraudsters and non-fraudsters in public transport who were or not exposed to ticket inspections by the transport company. We then measured the intrinsic honesty of the same persons in a new unrelated context where they could misappropriate money. Instead of having an educative effect across contexts, the exposure to deterrence practices increases unethical behavior of fraudsters but also of non-fraudsters.

Keywords:

Deterrence Institutions, Intrinsic Honesty, Spillovers, Field Experiment

JEL codes:
C93, K42, D02, D91

Fraud Deterrence Institutions Reduce Intrinsic Honesty

Fabio Galeotti[a], Valeria Maggian[b] and Marie Claire Villeval[c,d]

Abstract

Deterrence institutions are widely used in modern societies to discourage rule violations but whether they have an impact beyond their immediate scope of application is usually ignored. Using a natural field experiment, we show that they affect intrinsic honesty across contexts. We identified fraudsters and non-fraudsters in public transport who were or not exposed to ticket inspections by the transport company. We then measured the intrinsic honesty of the same persons in a new unrelated context where they could misappropriate money. Instead of having an educative effect across contexts, the exposure to deterrence practices increases unethical behavior of fraudsters but also of non-fraudsters.

JEL-Classification: C93, K42, D02, D91

Keywords: Deterrence Institutions, Intrinsic Honesty, Spillovers, Field Experiment.

[a] Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France. galeotti@gate.cnrs.fr

[b] Ca` Foscari University of Venice, Department of Economics. Cannaregio 873, Fondamenta San Giobbe, 30121 Venice, Italy. valeria.maggian@unive.it

[c] Corresponding author: Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France. villeval@gate.cnrs.fr

[d] IZA, Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany

Acknowledgments: We thank J. Andreoni, M. Bigoni, G. Charness, R. Kranton, J. Van de Ven, and audiences at seminars (at the Universities of Aarhus, Ancona, Bilkent, Birmingham, Innsbruck, Maastricht, Montpellier, Paris 1, Venice) and conferences (AFSE conference in Paris, ASFEE conference in Nice, 1st BEEN meeting at the University of Bologna, CBT Research day of EMLyon, ESA North-American meeting in Antigua, ESA World meeting in Berlin, EWEBE meeting in Tilburg, GATE-Lab-CORTEX workshop in Lyon, Organizations and Markets workshop in Dijon, Workshop on Social Economy for Young Economists in Bologna, 10th Birthday Conference of the Collegium of Lyon) for useful comments. We are grateful to R. Sauter and the TCL company in Lyon for facilitating our access to the transport network. This research has benefited from financial support from the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by Agence Nationale de la Recherche (ANR) and from the INDEPTH program from IDEXLYON (No 183634). This project has also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 661645. The IRB of INSERM (IRB00003888) has approved the project (decision n°17-416).

1. Introduction

Honesty and norm compliance are fundamental for the maintenance of trust and the development of prosperous societies (Mauro, 1995; Knack and Keefer, 1997). Norms can be sustained by an internalization mechanism that makes individuals comply even in absence of any threat of punishment (Axelrod, 1986; Gintis, 2003). However, intrinsic honesty is not sufficient to prevent violations and it varies widely across cultures (Fisman and Miguel, 2007; Cohn *et al.*, 2014; Gächter and Schulz, 2016). While peer punishment of violations facilitates compliance (Fehr and Gächter, 2000; Masclet *et al*., 2003), inspections and sanctions by centralized authorities are the most common institutional practices adopted in modern societies to deter deviant behavior. When they raise the costs of breaking the rule above its benefits, these institutions can discourage the targeted misbehavior (Becker, 1968; Di Tella and Schargrodsky, 2004; Fisman and Miguel, 2007; Baldassarri and Grossman, 2011; Ariely, 2012). However, they sometimes crowd out the intrinsic motivation to comply (Gneezy and Rustichini, 2000; Frey and Jegen, 2001; Falk and Kosfeld, 2006; Dickinson and Villeval, 2008), with potential spillovers onto adjacent activities (Belot and Schröder, 2016). This results from control averse individuals who directly reciprocate against a distrusting authority that reduces their freedom of choice.

While past estimations of deterrence effects focus almost exclusively on the targeted misbehavior, we defend that indirect effects may expand across contexts and impact both compliers and non-compliers. For example, it has been found that past exposure to institutions fostering pro-social norms can improve future pro-sociality even when the institution is no longer enforced (*e.g*., Cassar *et al.,* 2014; Peysakhovich and Rand, 2016; Engl *et al*., 2017; Galbiati *et al.,* 2018). Here, we look at possible spillover effects across contexts from inspecting and sanctioning people for rule violations on one of the most fundamental traits of human beings: intrinsic honesty. Investigating whether these spillovers exist is essential to better understand the overall effectiveness of these institutions, which crucially depends on whether they also affect socially desirable behavior beyond their immediate scope of intervention. However, to the best of our knowledge, no one has ever shed light on this issue.[1]

Why should we expect spillover effects of deterrence institutions on individuals' intrinsic honesty across contexts? The traditional economic approach to crime (Becker, 1968) is silent on their existence.[2] Alternatively, psychological and behavioral economics theories could account for these effects – which may be positive or negative. Focusing on dishonest individuals, if the experience of a deterrence institution in the past recalls what the society expects from individuals, this may act as an educative tool for the future and foster intrinsic honesty (on the socio-pedagogical effect of punishment see, *e.g.*, Hawkins, 1969; Andenaes, 1974; Hampton, 1984). At the same time, individuals who are caught breaking the rule are usually fined. In a subsequent unrelated context, they may be tempted to misbehave again in order to recover their financial loss (Sharma *et al.*, 2013). Intrinsic honesty may also decrease if dishonest individuals evaluate their moral activities dynamically (Nisan, 1991; Effron and Conway, 2015) and consider that the sanction has cleaned their past immoral actions ("*I paid for my sin*"), reducing the discrepancy between one's perceived self-image and the desired moral self.

Exposure to a deterrence institution may also have spillover effects on the intrinsic honesty of norm-compliers. On the one hand, the educative effect of deterrence institutions can act as a positive reinforcer. On the other hand, applying signalling theories (*e.g.*, Benabou and Tirole, 2003) could predict that an inspection reminds some people that their true intrinsic motivation for compliance is avoiding a fine, and crowd-out their intrinsic honesty in subsequent contexts where they know the deterrence institution is not anymore in place. Also, because of social learning, the

---

[1] The literature that identified negative effects of monitoring people on different productivity dimensions (*e.g.*, Gneezy and Rustichini, 2000; Falk and Kosfeld, 2006; Dickinson and Villeval, 2008) evaluated these effects only in the context where this institution directly operates. For example, Belot and Schröder (2016) show that controlling employees' performance may reduce their punctuality. These spillover effects within the same context are usually explained by direct reciprocity. But this literature ignores whether these effects spill over in other contexts that are not regulated by the institution – where direct reciprocity is ruled out – by affecting individual's intrinsic motivation.

[2] Most of the literature on deterrence in the Beckerian tradition examines whether variations in the probability of detection *vs*. severity of sanctions affect criminality (see Chalfrin and McCrary, 2017, for a survey). The only spillovers that are considered are those related to crime displacement following a sudden increase in the intensity of police (see review in Weisburd *et al.*, 2006), or those related to the incidence of more serious crimes following an increase in the intensity of arrest for small crimes (*e.g.*, Wilson and Kelling 1982). These studies – mostly conducted at the aggregate level – tend to be afflicted by simultaneity bias, omitted variables, and identification problems (Chalfrin and McCrary, 2017). In addition, they do not inform on spillover effects of the enforcement of the institution on intrinsic honesty. They only consider whether offenders reduce their criminal activities or relocate somewhere else after they update their perceived risk of apprehension in response to an increase in policing.

enforcement of the deterrence institution may affect compliers' beliefs about the spread of norm violation in society. Observing many violators being punished may reveal that misconduct is socially widespread and has become the norm, and lead compliers to behave accordingly (Keizer *et al*., 2008; Dickinson *et al*., 2015).

In sum, there exists several mechanisms that could lead to spillover effects of deterrence institutions across contexts at the individual level. However, whether these spillover effects truly exist in the real world, whether they are positive or negative, and whether they equally affect rule compliers and non-compliers remain open questions. To shed light on this, we ran a natural field experiment in public transport and on the streets in Lyon, France, with 708 passengers. We collected a direct and unbiased measure of dishonest behavior (*i.e*., fare evasion). Individuals were observed in a daily-life situation and were not aware of taking part in a study. Running a natural field experiment allows us to overcome all the limitations associated with laboratory experiments (Levitt and List, 2007; List, 2011) especially when investigating dishonesty, since the possible scrutiny by the experimenter can considerably affect unethical behavior (Gneezy *et al*., 2018). Moreover, eliminating any experimenter demand effect is of utmost importance to rule out direct reciprocity as a possible explanation of our results, especially when formal monitoring can be perceived as a form of distrust. Finally, besides contributing to the analysis of the dynamics of unethical behavior (but in different sense than Welsh, 2015, and Garrett *et al.,* 2016, who look at escalation effects), we focus on both compliers and non-compliers.

We chose to conduct our field experiment in public transport because in France all socio-demographic categories use public transport and fare evasion is relatively widespread (Cour des Comptes, 2016; Dai *et al*., 2018).[3] This means that when we study fraudsters, we are not looking at a tiny minority of people. Another reason is that it is a setting where dishonest behavior is perfectly and publicly identifiable with almost no measurement error since every passenger must validate a ticket or a pass every time he or she gets in a public vehicle.

The experiment consists of two stages. The first stage takes place on board buses and trams and produces two main natural conditions. In the *Inspection condition*, the targeted passenger has

[3] A 2011 survey conducted by OpinionWay in Lyon for the local public transport company revealed that 55% of the participants sometimes travel without a valid ticket (Keolis, 2014). The company estimates that around 1 out of 7 trips on the tram or bus is irregular (www.sytral.fr, accessed 02.08.2019).

been controlled by ticket inspectors from the transport company during his or her journey, whereas in the *No-Inspection condition* no ticket inspection occurred.[4] The second stage takes place when this targeted passenger gets off the vehicle, on the street. A professional actor who is part of the experimental team walks behind the targeted passenger and suddenly bends down to seemingly pick up a banknote on the ground. The actor then calls the attention of the targeted passenger by asking whether he or she had lost the banknote. We measure intrinsic honesty by looking at whether the passenger takes or not the banknote. We test whether this correlates with compliance in public transport. To identify the causal effect of the deterrence institution on compliers and non-compliers, we then contrast intrinsic honesty in the Inspection *vs*. No Inspection conditions.

We found that instead of having a positive immediate educative effect, the direct exposure to a deterrence institution in public transport increased unethical behavior of both fraudsters *and* non-fraudsters on the street. The effect was highly significant, it had the same magnitude for both groups (between 14% and 19% of the base level) and it increased with the size of the ticket inspection team. This shows that evaluating the full effect of deterrence institutions requests considering their spillover effects on intrinsic honesty, even in contexts where these institutions do not directly apply.

2. Experimental Design

Our field experiment consisted of two stages and was conducted by teams composed of a research assistant and a professional actor, both blind to the hypotheses of the study.[5,6] The first stage aimed at identifying dishonesty in a natural setting where formal deterrence institutions could be

---

[4] Note that our strategy of identification is not based on shocks in the deterrence policy but on natural variations in its implementation and on the observation of (dis-)honesty and exposure at the individual level. This implementation, we believe, acts as a reminder of the existence of the institution and, thus, as a proxy of a change in the institution.

[5] We used professional actors to ensure that the scene was played as similarly as possible across conditions. Four actors (two males, two females) were selected after a casting with 18 candidates. 21 subjects from the subject pool of the GATE-Lab in Lyon were recruited via Hroot (Bock *et al*., 2014) and paid €15 to evaluate the actors in terms of performance, honesty, trustworthiness, attractiveness, credibility, seriousness, and friendliness. We selected those actors with similar high scores in performance and credibility (see Figure A1 in Appendix 4). Before us, only few studies have used professional actors (Fischer *et al*., 2006; Swami *et al*., 2008; Gino *et al*., 2009; Wang *et al.,* 2012; Antonakis *et al*., 2015; Sands, 2017; Winter and Zhan, 2018).

[6] Note that 30% of the observations have been collected in the presence of an experimentalist who was not blind to the research questions, due to the unavailability of the assistant or because in the Audience condition – presented below – we needed two assistants (one to walk by the actor and another one collecting observable characteristics). As shown in the next section, this presence did not affect the results.

enforced. It took place in the public transport of Lyon (France) where the identification of fraudsters and non-fraudsters is direct: in order not to incur a fine, all passengers need to validate their ticket or pass at machines located on board public vehicles each time they enter a new vehicle. Re-validation is compulsory even if the ticket is valid for one hour from the first validation and has been already validated in a previous vehicle. Thus, someone who does not validate a pass or a ticket is in an irregular situation, and he or she is classified as a fraudster in our analysis.[7]

In the first stage of the experiment, the research assistant and the actor traveled on board a bus or tram. The former had to stay next to a validating machine and focus attention on the first four of five passengers boarding and validating or not their ticket. The actor waited on board the public vehicle without giving the impression of travelling with the research assistant. Once the first of these passengers got off the vehicle, both the research assistant and the actor also got off behind the targeted passenger. There were two conditions that occurred naturally. In the Inspection condition (I, hereafter), the targeted passenger was controlled by a team of ticket inspectors from the transport company during or at the end of the ride, whereas in the No-Inspection condition (NI, hereafter), no ticket inspection occurred.

The second stage of the experiment was conducted on the street, where we measured the intrinsic honesty of the same targeted passengers in a context where no formal institution applies. The actor, while having a fake phone conversation to minimize interactions, suddenly bent down to seemingly pick up a 5 euro banknote on the ground, just behind the targeted passenger. The actor then called the attention of the targeted passenger by asking whether he had lost the banknote. Accepting or not the banknote is our measure of intrinsic honesty. Meanwhile, the research assistant observed the scene and collected data on a tablet, regarding the decision to accept or not the banknote, any observable characteristics of the passenger (*e.g.*, apparent wealth and age, gender, emotional reaction to an inspection) and the environment (*e.g.*, approximate number of people on board, number of ticket inspectors, payment of a fine). The actor was instructed to play the scene with no audience at hearing distance. As a robustness check of the importance of audience on intrinsic honesty we ran an additional condition, the No-Inspection-Audience

[7] After scanning a ticket or a pass, the machine emits a clear beeping sound which makes forgetting to validate unlikely if other people are boarding at the same time. In buses, front door entry is compulsory but drivers have no responsibility in checking validation and they actually do not inspect. The only possible measurement error is when a passenger validates a ticket with a special tariff (*e.g.,* tariffs for seniors or unemployed) he or she is not entitled to.

condition (NI-A, hereafter). Here, the assistant walked by the actor and explicitly observed the scene. This allowed us to isolate the role that an observer plays in influencing individuals' unethical behavior.

In order to test that (*i*) one's false claim of ownership of a banknote found by someone else violates an injunctive ethical norm, and that (*ii*) the actors played the scene similarly across conditions, we conducted an additional laboratory experiment (see details about "Laboratory Experiment 1" in Appendix 2 and instructions in Appendix 3).[8] Specifically, using the elicitation procedure of Krupka and Weber (2013), we found that accepting a banknote found by another person was indeed considered as "very socially inappropriate" by 63% of the 30 subjects who participated in the lab experiment and no one considered it as "appropriate" (see Table A1 in Appendix 5).[9] To verify that the actors played the scene similarly across conditions, we recorded the actor's fake phone conversation during the field experiment, and 45 subjects in our laboratory experiment were incentivized to predict the decision of the targeted passenger after listening to 48 randomly selected audio files (12 for each main condition, I and NI, and each category of passengers, fraudsters and non-fraudsters).[10] Using the measure of guessing ability of Belot and van de Ven (2017),[11] we found that lab subjects were not able to predict the behavior of the field subjects. In all conditions and groups of field subjects, our measure of the ability to predict field behavior is either not significantly different from zero or weakly negative (in the I condition for fraudsters), meaning that lab subjects were, if anything, worse than chance in predicting field

---

[8] The laboratory experiment was fully computerized using the z-Tree software (Fischbacher, 2007), and conducted at GATE-Lab, Lyon (France). We recruited subjects with the online software Hroot (Bock *et al*., 2014). Subjects earned €5 for each correct guess of the modal response of the other participants.

[9] In contrast, claiming ownership of a banknote found on the ground by oneself was considered as somewhat or very socially appropriate by 80% of the subjects (see Appendix 2 and Table A1 in Appendix 5), which is significantly different from the answer in the other scenario (rank-sum test: $z = -4.84$, two-sided $p < 0.001$).

[10] For each audio file, subjects listened to the voice of the actor asking whether the targeted person has lost the €5 banknote but not the answer of the person. They were asked to guess whether or not the targeted person took the banknote. At the end, we drew 5 guesses for each participant and paid him or her €4 for each correct guess.

[11] The guessing ability is defined as $A = F(T \mid T) - F(T \mid NT)$, where $F(T \mid T)$ is the proportion of lab participants that guessed "took" when the field subject indeed took the banknote, while $F(T \mid NT)$ is the proportion of lab participants that guessed "took" when the field subject did not take the banknote. If $A \leq 0$, the lab subjects are not able to predict the behavior of the field subjects (*i.e.*, the probability of guessing that the person took the banknote is independent of the actual behavior of the person) or are worse than chance.

behavior (see Appendix 2 and Tables A2 and A3 in Appendix 5). We can therefore safely conclude that the actors played the scene similarly across conditions and with fraudsters and non-fraudsters.

Our experimental design, combining the three conditions described above (I, NI, NI-A) with the regular or irregular condition of the passenger on board the bus or tram, allows us to achieve a twofold objective: first, to investigate whether there is a correlation between the honesty of passengers in the bus/tram and on the street and, second, to identify the causal effect of ticket inspection on the latter.

In total, we collected 708 observations in the field from 358 non-fraudsters (104 in the I condition, 140 in NI and 114 in NI-A) and 350 fraudsters (100 in the I condition, 140 in NI, and 110 in NI-A).[12] We excluded vulnerable persons, minors and tourists (based on subjective judgment), persons accompanied by children, friends, colleagues or partner (see rules for subject selection in Appendix 1 and Table A4 in Appendix 5 for descriptive statistics on individual characteristics, showing that except for anticipated differences in apparent wealth between fraudsters and non-fraudsters, the sample is fairly balanced across conditions). The experiment was run in 2017. In a typical week day, we collected on average 21 observations over two time slots (between 9:00am and 12:30pm, and between 1:30pm and 5:30pm).

3. Results

Our results show that not validating the ticket in public transport is associated with a lower intrinsic honesty. In the absence of ticket inspection in the first stage, passengers without a validated ticket or pass (*i.e.*, fraudsters) were more likely than passengers with a validated ticket or pass (*i.e.*, non-fraudsters) to claim ownership of the banknote on the street in the second stage. Figure 1 presents the percentage of fraudsters and non-fraudsters who took the banknote in the NI and NI-A conditions. In NI, 52.86% of the fraudsters took the banknote compared to 32.14% of the non-fraudsters. This difference is significant (Chi-squared test: $\chi^2(1) = 12.29$, $p < 0.001$),[13] revealing that the difference in ethical behavior correlates across the two contexts. The observed pattern of cross-context unethical behavior is not affected by the presence of an observer in the second stage.

---

[12] The sample size was determined by following standard practices in the literature, using a significance level of 0.05 and setting power to 0.80. See Appendix 1 for details about the sample size calculation and the sample composition.

[13] All the reported non-parametric statistics are two tailed and take each individual as one independent observation.

Although fraudsters were slightly less likely to take the banknote when being observed by a third person (45.45% in NI-A *vs*. 52.86% in NI), the difference is not significant ($\chi^2(1) = 1.35$, $p = 0.245$). Similarly, the percentage of non-fraudsters who took the banknote in NI-A (33.33%) is not different from NI ($\chi^2(1) = 0.04$, $p = 0.841$).

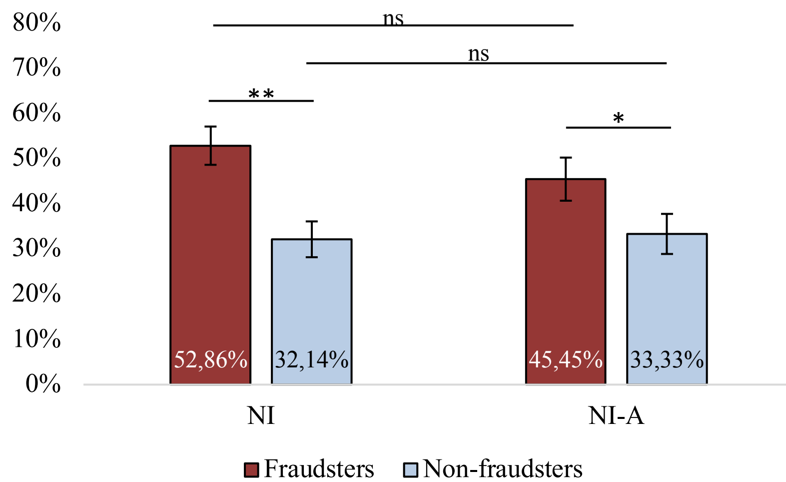


Figure 1. Percentage of fraudsters and non-fraudsters accepting the banknote in the conditions with no ticket inspection.

Notes: The light bars are for non-fraudsters and the dark bars for fraudsters in the two conditions without prior ticket inspection (NI and NI-A). N = 140 (NI, fraudsters), 140 (NI, non-fraudsters), 110 (NI-A, fraudsters), and 114 (NI-A, non-fraudsters). Error bars, mean ± SEM. *** $p < 0.01$, * $p < 0.1$, *ns* not significant, Chi-squared tests.

We found evidence that the enforcement of the deterrence institution in the first stage of the experiment spills over on the intrinsic honesty of fraudsters in the second stage (*i.e.*, behavior on the street). Figure 2 displays the percentage of fraudsters and non-fraudsters who took the banknote, depending on whether a ticket inspection occurred (I) or not (NI). This spillover effect is negative: the percentage of fraudsters who took the banknote increased significantly from 52.86% to 67% after an inspection ($\chi^2(1) = 4.81$, $p = 0.028$). This reveals that inspections and sanctions have no immediate educative effect on the intrinsic honesty of fraudsters.

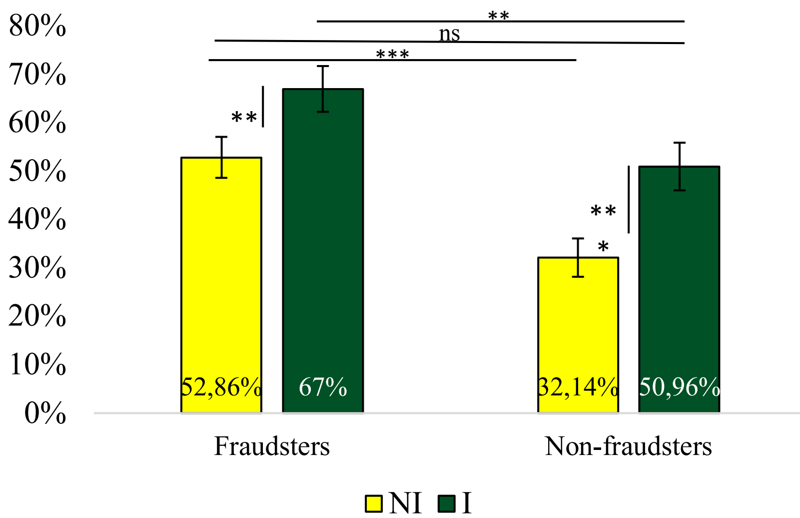


Figure 2. Percentage of fraudsters and non-fraudsters accepting the banknote in the conditions with (I) and without (NI) ticket inspection.

Notes: The light bars are for the NI condition and the dark bars for the I condition. N = 140 (NI, fraudsters), 100 (I, fraudsters), 140 (NI, non-fraudsters), and 104 (I, non-fraudsters). Error bars, mean ± SEM. *** $p < 0.01$, ** $p < 0.05$, *ns* not significant, Chi-squared tests.

Fraudsters who have been caught travelling irregularly had to pay a fine. Hence, the mechanism behind this negative spillover could be that fraudsters try to partly recover the loss caused by the fine (Sharma *et al.,* 2013). But more surprising is the behavior of non-fraudsters after a ticket inspection. If loss recovery was the only mechanism that generates the spillover, we should observe no spillover effect for non-fraudsters. Strikingly, the percentage of non-fraudsters accepting the banknote increased from 32.14% to 50.96% after ticket inspection ($\chi^2(1) = 8.79$, $p = 0.003$). The enforcement of the deterrence institution reduced the intrinsic honesty of the law-abiding passengers. The percentage of passengers who took the banknote after an inspection is still significantly higher for fraudsters than non-fraudsters ($\chi^2(1) = 5.41$, $p = 0.020$), but it no longer differs between non-fraudsters after an inspection and non-inspected fraudsters, even without an audience ($\chi^2(1) = 0.08$, $p = 0.769$).

Finally, the probability to take the banknote increased in the intensity of the inspection, as measured by the number of controllers conducting the ticket inspection, and this was observed for both fraudsters and non-fraudsters. Figure 3 displays the relationship between the number of inspectors and the probability to accept the banknote, as computed from the logit regression presented in Model 4 of Table 1. This analysis shows that it is not only the enforcement of the deterrence institution that affects intrinsic honesty across contexts, but also its intensity.

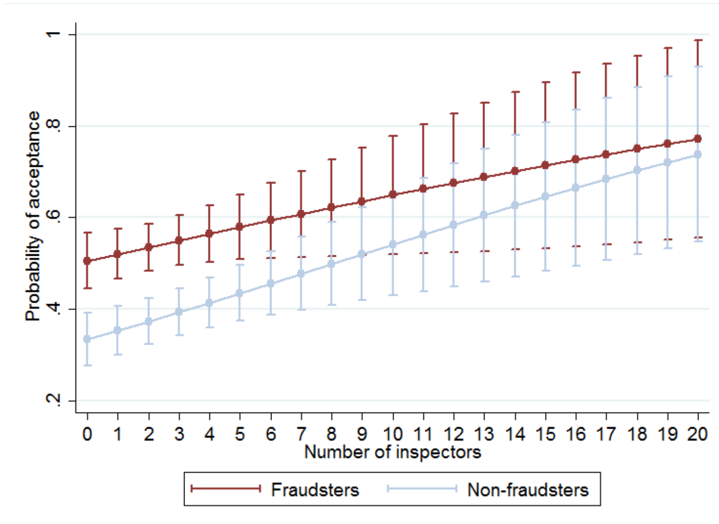


Figure 3. Probability to accept the banknote on the street depending on the intensity of ticket inspection in public transport.

Notes: The blue line is for fraudsters and the red line for non-fraudsters. The probability is computed from the Logit model (Model 4) reported in Table 1. N = 65 for [1, 5), 90 for [5, 10), and 45 for [10, 20]. Four observations were dropped because the information about the number of inspectors was missing.

We now turn to a regression analysis to control for the environment and for the individuals' socio-demographic characteristics. The average marginal effects from four Logit regressions in which the dependent binary variable is the decision to take or not the banknote are reported in Table 1. In Model 1, the independent variables only indicate whether the subject was or not a fraudster in public transport in each condition, with fraudsters in the NI condition taken as the baseline category. Model 2 augments this model with dummies for each actor with the high score actress (based on the evaluation by the subjects during the casting phase) taken as the baseline category, passengers' socio-demographics (apparent age, ethnicity, and wealth, gender, religious signs) and environmental conditions (type of vehicle, audience in the public vehicle, audience on the street, weather, main tram line, time of the day, geolocation). Model 3 is similar to Model 2 except that we replace the subject's gender dummy and the actor dummies with three indicator variables capturing the gender composition of the actor-passenger pair (with female pairs as the baseline category). Model 4 isolates the role of the number of ticket inspectors during a control. It is similar to Model 2, except that for identification we replace the variables indicating whether the subject was or not a fraudster in each condition with two dummies indicating whether the subject was a fraudster or not, regardless of the condition, and if there was an audience on the street (NI-A condition).

The general model that we fit is:

$$\mathrm{pr}(\text{TAKING THE BANKNOTE} = 1 \mid \boldsymbol{z}) = F(\beta_0$$

$$+ \beta_1 \text{ FRAUD}_{\text{i}} + \beta_2 \text{ NON-FRAUD}_{\text{ni}} + \beta_3 \text{ NON-FRAUD}_{\text{i}} + \beta_4 \text{ FRAUD}_{\text{a-ni}} + \beta_5 \text{ NON-FRAUD}_{\text{a-ni}} \tag{1}$$

$$+ \boldsymbol{\lambda} \textbf{ ACTOR} \tag{2a}$$

$$+ \boldsymbol{\varphi} \textbf{ DEMO} + \boldsymbol{\gamma} \textbf{ ENVIRON} \tag{2b}$$

$$+ \boldsymbol{\theta} \textbf{ GENDER\_PAIR} \tag{3}$$

$$\left.\begin{array}{l} + \varphi \text{ NON-FRAUD} + \psi \text{ NI-A} + \vartheta\,(\text{NUM\_INSPECTORS} \times \text{NON-FRAUD}) \\ \qquad + \omega \text{ NUM\_INSPECTORS} \end{array}\right) \tag{4}$$

where F($\mathbf{z}$)=$e^{\mathbf{z}}/(1 + e^{\mathbf{z}})$ corresponds to the cumulative logistic distribution. The variables FRAUD and NON-FRAUD identify whether a subject was or not a fraudster (with the subscript identifying the experimental condition). The vector **ACTOR** includes dummies for each actor, while **DEMO** and **ENVIRON** are vectors of control variables capturing the passengers' socio-demographics and the environmental conditions, respectively. The vector **GENDER_PAIR** contains dummies capturing the gender composition of the actor-passenger pair. Finally, the variables NI-A and NUM_INSPECTORS identify the audience condition and the number of ticket inspectors during a control, respectively. Model 1 only includes (1), Model 2 includes (1), (2a) and (2b), Model 3 comprises (1), (2b) and (3) but without the gender of the passenger in **DEMO**, while Model 4 contains (2a), (2b) and (4).

Table 1. Determinants of the decision to take the banknote.

Dependent variable: Decision to take the banknote	Model (1)		Model (2)		Model (3)		Model (4)	
	dy/dx	S.E.	dy/dx	S.E.	dy/dx	S.E.	dy/dx	S.E.
Treatment (Baseline = Fraudster NI)								
Non-fraudster NI	−0.21***	0.06	−0.19***	0.06	−0.18***	0.06	-	-
Fraudster I	0.14**	0.06	0.14**	0.07	0.14**	0.07	-	-
Non-fraudster I	−0.02	0.06	−0.04	0.07	−0.02	0.07	-	-
Fraudster NI-A	−0.07	0.06	−0.05	0.07	−0.05	0.07	-	-
Non-fraudster NI-A	−0.20***	0.06	−0.13**	0.07	−0.15**	0.07	-	-
Actors/Actress (baseline = Higher−score actress)								
Lower-score actress	-	-	0.20***	0.05	-	-	0.19***	0.05
Higher-score actor	-	-	−0.02	0.07	-	-	−0.03	0.07
Lower-score actor	-	-	0.09*	0.05	-	-	0.07	0.05
Additional controls for passengers (ethnicity, gender, social appearance, religious signs)	No		Yes		Yes		Yes	
Additional controls for the environment (weather, bus, main line, time of the day, geolocation, audience in the public vehicle, audience on the street)	No		Yes		Yes		Yes	
Gender interaction (baseline = Female actress, Female passenger*)*								
Female actress, Male passenger	-	-	-	-	0.03	0.05	-	-
Male actor, Female passenger	-	-	-	-	−0.05	0.06	-	-
Male actor, Male passenger	-	-	-	-	−0.02	0.05	-	-
Non-fraudster *(baseline = fraudster)*	-	-	-	-	-	-	−0.16***	0.04
Number of ticket inspectors	-	-	-	-	-	-	0.02***	0.01
Audience *(baseline = no audience)*	-	-	-	-	-	-	−0.01	0.05
Number of observations	708		708		708		704[§]	
Prob > chi2	<0.001		<0.001		<0.001		<0.001	
Pseudo R2	0.041		0.111		0.090		0.110	

Notes: Table 1 reports the average marginal effects from Logit estimates. [e] estimated by the research assistant. [§] Four observations were dropped because the information about the number of inspectors was missing. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (Wald tests).

The differences-in-differences regression analysis confirms that the spillover effect is of the same magnitude for fraudsters and non-fraudsters and thus, loss recovery cannot be the only explanation of these cross-context spillover effects. Indeed, all regressions indicate that fraudsters are 14% more likely to take the banknote in I than in NI (χ^2 test, $p = 0.025$).[14] Also, in the NI condition, non-fraudsters are 21% less likely to take the banknote than fraudsters ($p < 0.001$). In contrast, there is no significant difference between fraudsters in the NI condition and non-fraudsters in the I condition ($p = 0.769$). Audience in the NI-A condition does not make a difference

[14] The reported tests are based on Model 1. The results are analogous for the other models.

compared to the no audience conditions (p = 0.244). While one actor was more convincing than the others (Model 2) the main effects of the conditions remain when we control for these individual characteristics, as well as for the gender composition of pairs (Model 3).[15] A few socio-demographics matter: older subjects are more likely to take the banknote (possibly driven by a selection effect, as on average, wealthier older people use less public transport), while people with a wealthier appearance are less likely to violate the norm. The environmental conditions only marginally affect behavior (the effect of audience in public transport is weakly significant in Model 2; the probability of accepting the banknote is weakly significantly smaller in the afternoon in Models 2 and 4). Finally, Model 4 confirms that the number of ticket inspectors during a control in public transportation increases the likelihood of the norm violation on the street (p = 0.002).

4. Discussion

Our results provide strong evidence for cross-context spillover effects of inspections and sanctions on intrinsic honesty. Strikingly, these effects equally apply to fraudsters and to non-fraudsters. In what follows, we examine some existing mechanisms that could explain these spillover effects.

Negative *direct reciprocity* against the authority that signals distrust by enforcing inspections is ruled out by design since behavior on the street cannot affect the transport company. But people may still want to harm a stranger because of *indirect reciprocity* (Nowak and Sigmund, 2005). While we cannot exclude this possibility, it seems unlikely for a number of reasons. First, it is unclear why an inspected passenger would like to hurt a 'kind' third party who has just offered him or her money. Second, indirect reciprocity often arises for strategic motives (*e.g.*, Engelman and Fischbacher, 2009; Stanca, 2009) that are absent in our setting.

Fraudsters' willingness to *recover a loss* might be a good candidate to explain their subsequent unethical conduct. However, it cannot explain the negative effect of inspections on non-fraudsters' intrinsic honesty across contexts. It is still possible that spillovers are caused by negative emotions that lead the passenger to punish whoever can be associated with the transport company (*e.g.*, another passenger) or society in general. In fact, we also reject explanations based on *emotions*

[15] As already mentioned, 30% of the observations have been collected in the presence of an experimentalist who was not blind to the research questions. If we directly control for this presence in the regression analysis the results do not change (see Table A5 in Appendix 5).

(see, *e.g.*, Card and Dahl, 2011; Munyo and Rossi, 2013). While anger for being fined might deplete the cognitive resources necessary to resist temptation, the banknote acceptance rate does not depend on whether detected fraudsters expressed a violent emotion during the inspection (68.75%) or not (67.57%) ($\chi^2(1) = 0.01$, $p = 0.927$). Moreover, a survey that we conducted with passengers in public transport several weeks after the experiment reveals that non-fraudsters' self-reported happiness and nervousness after a ticket inspection (N = 51, mean = 3.88 and 2.24, respectively, on a scale from 1 to 5) and when no inspection occurred (N = 109, mean = 4.14, and 1.93, respectively) do not differ significantly (Mann-Whitney test, z = 1.13, $p = 0.257$ for happiness; z = −1.149, $p = 0.251$ for nervousness).[16]

The experience of inspections may affect people's perception of the *injunctive norms* (what one ought to do or not to do) or the *descriptive norms* (what most people do), and this might affect behavior in a subsequent unrelated situation. Complementary investigations do not support this normative explanation. Laboratory Experiment 1 showed that claiming ownership of a banknote found on the ground by oneself was considered as "somewhat or very socially appropriate" by 80% of the subjects, whereas taking the banknote when someone else has found it was judged as "somewhat or very socially inappropriate" by 100% of them. In Laboratory Experiment 2, we elicited again injunctive norms following the same procedure as in Experiment 1 but *after* new subjects (N = 96) played a simplified version of the public transport game of Dai *et al.* (2017) (see details in Appendix 2 and instructions in Appendix 3). In this game, subjects had to decide whether purchasing or not a ticket, being uninformed of the exact probability of a control (50%). We reject that the injunctive norm differs between inspected and non-inspected non-fraudsters in this game.[17] However, inspections might still inform people on the descriptive norm in the field (Sliwka, 2007; Dickinson *et al.*, 2015). If ticket inspectors signal the prevalence of rule violations, after an inspection people may revise downward their perception of the descriptive norm in the society.

[16] We interviewed 109 non-fraudsters who had not been inspected and 51 non-fraudsters who had been inspected, following the same identification procedure as in our field experiment. We used Self-Assessment Manikins (SAM) (Lang, 1980) to elicit happiness and nervousness (see Figure A4 in Appendix 4). We also elicited passengers' beliefs about the percentage of fare evaders in the transport network in Lyon. See details of the procedures in Appendix 2.

[17] Claiming ownership of a banknote found on the ground by oneself was considered as "somewhat or very socially appropriate" by 93.54% of the non-inspected non-fraudsters in the game and by 88% of the inspected non-fraudsters (Mann-Whitney tests, $p = 0.823$). Claiming ownership when the banknote has been found by another person was considered as "somewhat or very socially inappropriate" by 96.78% of the non-inspected non-fraudsters in the game and by 92% of the inspected non-fraudsters ($p = 0.816$). See Table A6 in Appendix 5.

This may weaken their own morals and push them to accept undeserved money in a subsequent context. Our survey conducted in public transport contradicts this hypothesis. Its results show that ticket inspection does not change the beliefs of non-fraudsters in the field about the prevalence of fare evasion (non-inspected non-fraudsters: N = 108, mean belief about the percentage of fraudsters = 30.93%; inspected non-fraudsters: N = 50, mean belief = 29.28%. Mann-Whitney test, $z = 0.441$, $p = 0.659$. The result holds if we look at the number of controllers conducting the ticket inspection).

A psychological explanation in terms of *moral licensing* (Nissan, 1991) could apply to non-fraudsters if after a ticket inspection that reinforced their positive self-image, they loosen their moral standards while maintaining their self-concept of honesty (Benabou and Tirole, 2006; Shalvi *et al*., 2011). Symmetrically for fraudsters, paying a fine in itself may lead to *moral cleansing* if the sanction reduces the dissonance between the individual's self-image and his desired moral self. However, the acceptance rate of the banknote in our field experiment did not differ between the fraudsters who paid their fine on the spot (63.5%) – those for which such moral cleansing applies – and those who did not (73.2%) ($\chi^2(1) = 0.989$, $p = 0.320$).

Finally, an inspection might act on people as *a signal or a reminder* of their true nature (Benabou and Tirole, 2003). Not only fraudsters are reminded that they are dishonest, but a fraction of non-fraudsters are reminded that they comply by validating their ticket or pass just to avoid a sanction, and that without such deterrence institution they would evade fares. For these individuals, inspections may reinforce compliance when they know that the institution is enforced, but crowd out intrinsic honesty when they know it cannot be enforced.

Our findings are consistent with a number of possible mechanisms. While some of them are more plausible than others (in particular, the signaling interpretation), we cannot unambiguously isolate a unique explanation of the observed spillover effects. Also, the mechanism may work differently for fraudsters and non-fraudsters, though the size of the effect is analogous between the two groups of passengers. Note that our main objective was not to pin down the precise mechanism behind these spillover effects but to establish whether they exist in the first place and define their direction – something that was largely ignored in the literature. We discussed and shortlisted a number of possible explanations. Future studies can be devoted to identify what the main driver is among these alternative explanations.

5. Conclusion

Modern societies have developed centralized institutions to protect citizens and assets against dishonesty. Since the honesty norms prevailing in the environment, *i.e.,* the frequency of violations, can compromise intrinsic honesty in a society (Gächter and Schulz, 2016), one might expect that these institutions contribute to the elevation of intrinsic honesty. However, solely focusing on the impact of these deterrence institutions in their context of application does not permit to isolate their pure effect on intrinsic honesty since this is confounded by other factors such as simple material cost-benefit considerations (*e.g*., avoiding a sanction) or direct reciprocity. By identifying their causal effect outside their scope of application, our field experiment shows that the relationship between deterrence institutions and intrinsic honesty is more complex than what one might expect. Deterrence institutions create incentives to behave honestly to avoid a sanction but, at the same time, we have shown that they reduce intrinsic honesty. Instead of observing an educative effect across contexts, we found that inspections can lead even those who abided by the law to behave unethically in domains where the institution does not apply. This invites to adopt a broader view when evaluating the effectiveness of an institution. A social welfare perspective requests to make sure that, in the aggregate, the positive effects of an institution are not cancelled out by spillovers in contexts beyond its direct target.

References

Andenaes, Johannes. 1974. "*Punishment and Deterrence*." Michigan University Press.

Antonakis, John, Giovanna d'Adda, Roberto Weber, and Christian Zehnder. 2015. ""Just words? Just speeches?" On the economic value of charismatic leadership." *Working Paper Department of Organizational Behavior,* University of Lausanne.

Ariely, Dan. 2012. "*The honest truth about dishonesty. How we lie to everyone---especially ourselves*." New York: Harper Collins.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review,* 80:1095-1011. DOI: https://doi.org/10.1017/S0003055400185016

Baldassarri, Delia, and Guy Grossman. 2011. "Centralized sanctioning and legitimate authority promote cooperation in humans." *Proceedings of the National Academy of Sciences, 108(27):* 11023-11027. DOI: 10.1073/pnas.1105456108

Becker, Gary S. 1968. "Crime and punishment: an economic approach." *Journal of Political Economy,* 76(2): 169-217. DOI: https://doi.org/10.1086/259394

Belot, Michele, and Marina Schröder. 2016. "The spillover effects of monitoring: A field experiment." *Management Science,* 62(1):37-45. http://dx.doi.org/10.1287/mnsc.2014.2089

Belot, Michele, and Jeroen van de Ven. 2017. "How private is private information? The ability to spot deception in an economic game." *Experimental Economics,* 20: 19-43. DOI: 10.1007/s10683-015-9474-8

Benabou, Roland, and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *Review of Economic Studies,* 70(3): 489-520. DOI: https://doi.org/10.1111/1467-937X.00253

Benabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review,* 96(5): 1652-1678. DOI: 10.1257/aer.96.5.1652

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "Hroot: Hamburg Registration and Organization Online Tool." *European Economic Review,* 71(C): 117-120. DOI: 10.1016/j.euroecorev.2014.07.003

Card, David, and Gordon B. Dahl. 2011. "Family violence and football: The effect of unexpected emotional cues on violent behavior." *The Quarterly Journal of Economics*, 126 (1): 103-143. DOI: https://doi.org/10.1093/qje/qjr001

Cassar, Alessandra, Giovanna d'Adda, and Pauline Grosjean. 2014. "Institutional Quality, Culture, and Norms of Cooperation: Evidence from Behavioral Field Experiments." *Journal of Law and Economics,* 57(3): 821-863. DOI: http://dx.doi.org/10.1086/678331

Chalfin, Aaron, and Justin McCrary. 2017. Criminal Deterrence: A Review of the Literature. *Journal of Economic Literature*, 55(1): 5-48. DOI: https://doi.org/10.1257/jel.20141147

Cohn, Alain, Ernst Fehr, and Michel-André Maréchal. 2014. "Business culture and dishonesty in the banking industry." *Nature,* 516: 86-89. DOI: 10.1038/nature13977

Cour des Comptes. 2016. "La lutte contre la fraude dans les transports urbains en Île-de-France: un échec collectif. " *Rapport Public Annuel*, 537-577.

Dai, Zhixin, Fabio Galeotti, and Marie Claire Villeval. 2017. "The efficiency of crackdowns: a lab-in-the-field experiment in public transportations." *Theory and Decision,* 82(2): 249-271. DOI: https://doi.org/10.1007/s11238-016-9561-0

Dai, Zhixin, Fabio Galeotti, and Marie Claire Villeval. 2018. "Cheating in the lab predicts fraud in the field. An experiment in public transportations." *Management Science,* 64(3): 1081-1100. DOI: 10.1287/mnsc.2016.2616

Di Tella, Rafael, and Ernesto Schargrodsky. 2004. "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack." *American Economic Review,* 94(1): 115-133. DOI: 10.1257/000282804322970733

Dickinson, David L., E. Glenn Dutcher, and Cortney S. Rodet. 2015. "Observed punishment spillover effects: a laboratory investigation of behavior in a social dilemma." *Experimental Economics,* 18: 136-153. DOI: 10.1016/j.jpubeco.2010.11.021

Dickinson, David L., and Marie Claire Villeval. 2008. "Does monitoring decrease work effort? The complementarity between agency and crowding-out theories." *Games and Economic Behavior,* 63: 56-76. DOI: https://doi.org/10.1016/j.geb.2007.08.004

Effron, Daniel A., and Paul Conway. 2015. "When virtue leads to villainy: advances in research on moral self-licensing." *Current Opinion in Psychology,* 6: 32-35. DOI: https://doi.org/10.1016/j.copsyc.2015.03.017

Engelman, Dirk, and Urs Fichbacher. 2009. "Indirect reciprocity and strategic reputation building in an experimental helping game." *Games and Economic Behavior*, 67(2): 399-407. DOI: https://doi.org/10.1016/j.geb.2008.12.006

Engl, Florian, Arno Riedl, and Roberto A. Weber. 2017. "Spillover Effects of Institutions on Cooperative Behavior, Preferences, and Beliefs." *CESifo Working Paper No6504*, Munich (2017). Retrieved from http://www.cesifo-group.de/DocDL/cesifo1_wp6504.pdf on 3 August 2019.

Falk, Armin, and Michael Kosfeld. 2006. "The hidden costs of control." *American Economic Review,* 96(5): 1611-1630. DOI: 10.1257/aer.96.5.1611

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and punishment in public goods experiments." *American Economic Review,* 90(4): 980-994. DOI: 10.1257/aer.90.4.980

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics,* 10(2): 171-178. DOI: https://doi.org/10.1007/s10683-006-9159-4

Frey, Bruno S., and Reto Jegen. 2001. "Motivation crowding theory." *Journal of Economic Surveys,* 15: 589-611. DOI: https://doi.org/10.1111/1467-6419.00150

Fischer, Peter, Tobias Greitemeyer, Fabian Pollozek, and Dieter Frey. 2006. "The unresponsive bystander: Are bystanders more responsive in dangerous emergencies?" *European Journal of Social Psychology,* 36: 267-278. DOI: https://doi.org/10.1002/ejsp.297

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets." *Journal of Political Economy,* 115(6): 1020-1048. DOI: https://doi.org/10.1086/527495

Gächter, Simon, and Jonathan F. Schulz. 2016. "Intrinsic honesty and the prevalence of rule violations across societies." *Nature,* 531: 496-499. DOI: 10.1038/nature17160

Galbiati, Roberto, Emeric Henry, and Nicolas Jacquemet. 2018. "Dynamic effects of enforcement on cooperation." *Proceedings of the National Academy of Sciences,* 115(49): 12425-12428. DOI: https://doi.org/10.1016/j.jempfin.2018.07.008

Garrett, Neill, Stephanie Lazzaro, Dan Ariely, and Tal Sharot. 2016. "The brain adapts to dishonesty." *Nature Neuroscience* 19(12):1727-1732. DOI: 10.1038/nn.4426

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2009. "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel." *Psychological Science,* 20(3): 393-398. DOI: https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gintis, Herbert. 2003. "The hitchhiker's guide to altruism: Gene-culture coevolution and the internalization of norms." *Journal of Theorical Biology,* 220(4): 407-418. DOI: https://doi.org/10.1006/jtbi.2003.3104

Gneezy, Uri, Agne Kajackaite, and Joel Sobel. 2018. "Lying Aversion and the Size of the Lie." *American Economic Review,* 108(2): 419-453. DOI: https://doi.org/10.1257/aer.20161553

Gneezy, Uri, and Aldo Rustichini. 2000. "Pay enough or don't pay at all." *Quarterly Journal of Economics,* 115(3): 791-810. DOI: https://doi.org/10.1162/003355300554917

Hampton, Jean. 1984. "The moral education theory of punishment." *Philosophy and Public Affairs,* 13, 208-238.

Hawkins, Gordon. 1969. "Punishment and Deterrence: The Educative, Moralizing, and Habituative Effects." *Wisconsin Law Review* 550.

Keolis. 2014. "Fraude: comment lutter?" Available online at https://www.slideshare.net/Keolis/keo-fevrier-2014 (accessed 2 August 2019).

Knack, Stephen, and Philip Keefer. 1997. "Does social capital have an economic payoff? A cross-country investigation." *Quarterly Journal of Economics,* 112(4): 1251-1288. DOI: https://doi.org/10.1162/003355300555475

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association* 11(3): 495-524. DOI: https://doi.org/10.1111/jeea.12006.

Lang, Peter. 1980. "Behavioral treatment and bio-behavioral assessment: computer applications." in Sidowski J.B., J.H. Johnson, and T.A. Williams, eds. *Technology in mental health care delivery systems*. Norwood: Ablex, 119-l37.

Levitt, Steven D., and John A. List. 2007. "What do laboratory experiments measuring social preferences tell us about the real world?" *Journal of Economic Perspectives,* 21(2): 153-174. DOI: 10.1257/jep.21.2.153

List, John A. 2011. "Why economists should conduct field experiments and 14 tips for pulling one off." *Journal of Economic Perspectives,* 25(3): 3-16. DOI: 10.1257/jep.21.2.153

Masclet, David, Charles Noussair, Steven Tucker, and Marie Claire Villeval. 2003. "Monetary and non-monetary punishment in the voluntary contributions mechanism." *American Economic Review,* 93: 366-380. DOI: 10.1257/000282803321455359

Mauro, Paulo. 1995. "Corruption and Growth." *Quarterly Journal of Economics,* 110(3): 681-712. DOI: https://doi.org/10.2307/2946696

Munyo, Ignacio and Martín A. Rossi. 2013. "Frustration, euphoria, and violent crime." *Journal of Economic Behavior & Organization*, 89: 136-142. DOI: https://doi.org/10.1016/j.jebo.2013.02.005

Nisan, Mordechai. 1991. "The moral balance model: Theory and research extending our understanding of moral choice and deviation." in Kurtine, W.M., and J.L. Gewirtz, eds. *Handbook of Moral Behavior and Development,* Lawrence Erlbaum Associates, Inc., 3-213.

Nowak, Martin A., and Karl Sigmund. 2005. "Evolution of Indirect Reciprocity." *Nature*, 437(7063): 1291-1298. DOI: 10.1038/nature04131

Peysakhovich, Alexander, and David G. Rand. 2016. "Habits of virtue: creating norms of cooperation and defection in the laboratory." *Management Science,* 62(3): 631-647. DOI: http://dx.doi.org/10.1287/mnsc.2015.2168

Rand, David G., Alexander Peysakhovich, Gordon T. Kraft-Todd, George E. Newman, Owen Wurzbacher, Martin A. Nowak, and Joshua D. Greene. 2014. "Social heuristics shape intuitive cooperation." *Nature Communications,* 5: 3677. DOI: 10.1038/ncomms4677

Sands, Melissa L. 2017. "Exposure to inequality affects support for redistribution." *Proceedings of the National Academy of Sciences,* 114(4): 663-668. DOI: 10.1073/pnas.1615010113.

Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten K.W. De Dreu. 2011. "Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior." *Organizational Behavior and Human Decision Processes,* 115: 181-190. DOI:10.1016/j.obhdp.2011.02.001

Sharma, Eesha, Nina Mazar, Adam L. Alter, and Dan Ariely. 2014. "Financial deprivation selectively shifts moral standards and compromises moral decisions." *Organizational Behavior and Human Decision Processes,* 123(2): 90-100. DOI: http://dx.doi.org/10.1016/j.obhdp.2013.09.001

Sliwka, Dirk. 2007. "Trust as a signal of a social norm and the hidden costs of incentive schemes." *American Economic Review,* 97(3): 999-1012. DOI: 10.1257/aer.97.3.999

Stanca, Luca. 2009. "Measuring indirect reciprocity: Whose back do we scratch?" *Journal of Economic Psychology,* 30(2): 190-202. DOI: https://doi.org/10.1016/j.joep.2008.07.010

Swami, Viren, Flora Chan, Vivien Wong, Adrian Furnham, and Martin J. Tovée. 2008. "Weight-based discrimination in occupational hiring and helping behavior." *Journal of Applied Social Psychology,* 38: 968-981. DOI: https://doi.org/10.1111/j.1559-1816.2008.00334.x

Wang, Lu, Gregory B. Northcraft, and Gerben A. Van Kleef. 2012. "Beyond negotiated outcomes: The hidden costs of anger expression in dyadic negotiation". *Organizational Behavior and Human Decision Processes,* 119(1): 54-63. DOI: 10.1016/j.obhdp.2012.05.002

Weisburd, David, Laura A. Wyckoff, Justin Ready, John E. Eck, Joshua C. Hinkle, and Frank Gajewski. 2006. "Does Crime Just Move around the Corner? A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits." *Criminology,* 44(3): 549–592. DOI: https://doi.org/10.1111/j.1745-9125.2006.00057.x

Welsh, David, Lisa D. Ordóñez, Deirdre G. Snyder, and Michael S. Christian. 2015. "The slippery slope: how small ethical transgressions pave the way for larger future transgressions." *Journal of Applied Psychol*ogy, 100: 114-127. DOI: 10.1037/a0036950

Wilson, James Q., and George L. Kelling. 1982. "Broken Windows: The Police and Neighborhood Safety." *Atlantic Monthly*, 249(3): 29-38.

Winter, Fabian, and Nan Zhan. 2018. "Social norm enforcement in ethnically diverse communities." *Proceedings of the National Academy of Sciences,* 115(11): 2722-2727. DOI: 10.1073/pnas.1718309115.

APPENDIX 1: ADDITIONAL INFORMATION ON THE FIELD SETTING AND THE PROTOCOL OF THE FIELD EXPERIMENT

1.1. INTERNAL CONTROL OF THE PROTOCOL

Our field experiment was run in 2017 over many days (excluding week-ends). During this period the transport company did not introduce any particular change in its ticket inspection policy. In a typical day, we collected on average 21 observations over two time slots (between 9:00am and 12:30pm, and between 1:30pm and 5:30pm). We avoided rush hours because passengers may anticipate that the risk of ticket inspection is lower during these hours. To maximize the internal control of our experiment in the field, we proceeded as follows.

First, we ran the field experiment using professional actors who were blind to the research objectives. The professional actors were selected via a three-step procedure. We started by organizing a casting with 18 candidates in a professional acting school in Lyon. The candidates were former graduates of the school and students in their last year. Each candidate was asked to play the scene that we later used in the field experiment. We told the candidates only the details of the scene without revealing any information about our experimental objectives. They also played a second scene not related to our experiment (*i.e.*, asking a person whether they could borrow her/his mobile phone to make a call) to obfuscate our objectives. The candidates' performance was video recorded.

We then recruited 21 subjects from the subject pool of the GATE-LAB in Lyon via Hroot (Bock *et al*., 2014) to evaluate the actors. We paid subjects €15 for their participation. We asked them to watch the videos of the 18 actors and rate each candidate in terms of performance, honesty, trustworthiness, attractiveness, credibility, seriousness, and friendliness. We finally selected two actors and two actresses with similar high scores in performance and credibility, and similar scores in the other dimensions (one pair made of one actor and one actress with high scores in all the other dimensions, and another pair with lower scores) (see Figure A1 in Appendix 3).

Second, we asked the research assistant (also blind to the research objectives) to select the first four or five persons entering the bus or tram and observe whether they validated or not a ticket or a pass. The first of these persons exiting the vehicle became the target subject of the experiment. The research assistant was instructed to collect several information for each observation and entered them on a tablet. He recorded the name of the actor playing the scene, the time of the day, the weather (sunny, cloudy or rainy), the treatment condition (Inspection, No Inspection, No Inspection-Audience), the name of the bus/tram line where the subject travelled, whether the subject validated a ticket, a monthly pass or nothing, the name of the bus/tram stop where the subject got off the public vehicle, the approximate number of people on board the bus/tram (almost empty, quite crowded but everyone could sit, crowded), whether someone could notice the scene played in the street, whether the subject took or not the 5-euro banknote, the gender, estimated age (18-24, 25-34, 35-44, 45-59, 60 or more), estimated economic status based on appearance (poor, average, wealthy), and ethnicity (Caucasian, Arab, African, Asian, other) of the targeted passenger, and whether the subject wore religious symbols (*e.g.*, veil, kippah, showy Christian cross). In the Inspection condition, the research assistant also recorded the number of ticket inspectors, whether the inspection was conducted at the tram/bus stop or on board, whether the ticket inspectors wore uniform or civil clothes, the gender of the controller who inspected the targeted passenger, and, if applicable, whether the passenger paid the fine immediately, and whether he or she had an emotional or aggressive reaction during the control.

Third, while the actor was passive in the first stage of the experiment on board the bus/tram, he or she played a well-defined role in the second stage. We asked the actors to use their mobile phone as an audio recording device when playing the scene on the street.[18] We used these recordings to (*i*) verify that the actors played the scene according to the protocol, and (*ii*) as a robustness check to ensure that any minimal deviation from this protocol did not affect the internal validity of our results. We reproduce below (translated from French), the script given to the actors.

<u>Scene to play</u>

The actor/actress is on the phone. He/she follows a passenger, indicated by the RA, at the exit of the bus/tram. After 20/30 meters, he/she catches up and pretends to pick up a 5-Euro banknote from the floor. The actor/actress calls the attention of the target passenger, interrupting for a moment his/her phone call, to ask if the person has lost the banknote. The banknote must be clearly visible: it must be immediately clear that the banknote has just been picked up and that it is a 5-Euro banknote. The sentence to say is: "Sir/Madam, did you lose this?", in a neutral tone.

The actor/actress pretends not to pay too much attention to the targeted passenger's response and remains focused on the phone call he/she is having. If the passenger responds affirmatively, the actor/actress gives the banknote to the person without showing any signs of surprise. If the passenger responds negatively, the actor/actress puts the banknote in his/her pocket. In both cases, after the interaction, the actor/actress resumes his telephone call.

In the event that the passenger interacts with the actor/actress, for example by asking him/her if he/she has seen the money falling from the passenger's pocket, the actor/actress must avoid initiating a conversation. In the example just given, the actor/actress will simply answer "No idea" or "I did not pay attention".

1.2. THE PUBLIC TRANSPORT NETWORK IN LYON

The experiment was run in the main tram and bus lines of Lyon and we informed the transport company about our experiment. The public transport network in Lyon comprises 4 metro lines, 5 tramway lines, 2 funicular lines and over 130 bus lines. Tickets can be purchased from vending machines located at each tram stop and metro entrance, from dedicated agencies or from the bus driver on board the public vehicle. In 2017, when the experiment was run, a single ticket costed €1.80 at vending machines and €2 on board. The single ticket enables passengers to use the public transport network for an unlimited number of times and any distance during one hour. Passes require the client to buy a smartcard at a cost of €5 and then a pass which is held on the smartcard. In 2017, the monthly pass costed €63.20 and the yearly pass €60.10 per month, with a discount for people less than 21 years old and half of it being reimbursed by the employer, according to the labor law.

To avoid a fine, passengers must validate their ticket or pass every time they board a new public vehicle even if they have already validated it in a previous journey. The fine amounts to €60 if paid on the spot, while it increases to €80 (€110) if paid with a maximum delay of 7 days (2 months). If a person did not validate a pass, this is also considered as an infraction but the amount of the fine is reduced to €5.

Fare dodging is quite frequent in Lyon. A 2011 survey conducted by OpinionWay in Lyon for the public transport company revealed that 55% of the participants sometimes travel without a valid ticket (Keolis, 2014). The company also estimates that around 1 out of 7 trips on the tram or bus is irregular (www.sytral.fr, accessed 2.08.2019). This means that, in our Inspection condition, every passenger can, on average, observe at least one person getting a fine. This also means that we are not observing the behavior of a small minority of people.

---

[18] We thank James Andreoni for suggesting this to us. We did not record the audio of the scenes in the first experimental sessions. Also, due to some technical or environmental problems (*e.g.,* the actor forgot to press the record button or the quality of the audio was too poor), we failed to record the audio of few other scenes.

1.3. TICKET INSPECTIONS

The transport company conducts ticket inspections every day. Most inspectors wear official uniforms but some of them wear civil clothes. Ticket inspectors work in teams of different sizes (typically of 4) and they can be as many as 20 during a hot-spot inspection (this can happen if different teams meet in the transport network and decide to conduct more systematic inspections). Figure A2 in Appendix 4 shows a histogram with the distribution of the number of ticket inspectors per inspection that we encountered in our field experiment. The teams can be on board, changing line and direction as they wish, or waiting for the public vehicle to arrive at a stop. In the first case, the inspection is conducted during the ride. In the second case, it is done at the stop. In both cases, everyone on board is controlled. The inspector scans the client's ticket or pass in a device. If a passenger is caught fare dodging, the controller issues a fine. The enforcement of a fine takes several minutes (sometimes more if the passenger try to find excuses or confront the inspector). As a result, everyone is able to see a person who gets a fine. Ticket inspectors are paid a flat wage and receive no incentives for the number of fare-dodgers they catch or people they control.

According to the company, ticket inspections are mostly random and irregular in order to maintain uncertainty and prevent fare-dodgers to learn where inspections could be. The inspection plans change every day and are subject to unexpected changes within the day. Since inspectors have usually a large discretion regarding where to go within a predefined area, it is very difficult to localize them. We also checked whether there existed apps for smart phones able to signal the presence of inspectors in the public transport network. We did not find any that worked during the realization of our experiment. All this largely explains why we could not collect more than 21 observations per day on average.

Of course, ticket inspections might occur more frequently in certain lines or areas for logistical reasons (*e.g.,* accessibility of the zone, shift work organization, number of people using a line). We can check for this by looking at the frequency of inspections and their geolocation observed in our field experiment. We covered three main areas of Metropolitan Lyon: Center Metropolitan Lyon, Nord-East Metropolitan Lyon and South-East Lyon. These are areas that can be easily accessed by metro and tram. We did not cover West Metropolitan Lyon since it is a hilly area with no metro or tram stops. For logistical reasons, we also did not visit far-away neighborhoods in the East or South of Lyon. Figure A3 in Appendix 4 plots the frequency of inspections onto a map covering the area of Metropolitan Lyon which can be reached by tram or metro. Inspections were more frequent in Center Metropolitan Lyon (darkest blue shaded area) than Nord-East Metropolitan Lyon (medium blue shaded area; χ^2 test, $p = 0.029$) and South-East Metropolitan Lyon (light blue shaded area; $p = 0.002$). Nord-East and South-East Metropolitan Lyon present similar frequencies of inspections ($p = 0.294$). This is not surprising given that more lines pass through Center Metropolitan Lyon and it is visited by more people.

Turning to the tram and bus lines that we covered in our field experiment and focusing on the ones we visited the most (tram lines T1, T2 and T4),[19] we find that inspections were more frequent in the tram line T4 (33.33%) compared to any other lines (20.36% in T1 and 20.18% in T2; χ^2 tests, $p < 0.05$ for both T1 *vs*. T4 and T2 *vs*. T4). Line T4 vertically crosses the metropole of Lyon, and stops at two main train stations and the University Campus. So, it is a line that might require additional attention. Line T4 is called "main line" in Table 1 reported in the text.

We checked with the transport company the consistency between the frequency of inspections observed in our field data and those reflected in the inspection plans of the company for the periods corresponding to the experiment. Overall, this consistency was high and this confirms that our randomization strategy worked. In the econometric analysis reported in the article, we directly control for these aspects.

[19] We also travelled on a very wide range of bus lines but this was done less regularly, and typically when the public transport company or other controllers tipped us off about an inspection on a specific bus line.

1.4. SAMPLE SIZE AND SUBJECT SELECTION

The study involved 708 passengers. To determine the sample size for both the NI and I conditions, we conducted an *a priori* power analysis. To form reasonable predictions about the behavior of fraudsters and non-fraudsters in the control group (*i.e.*, NI condition), we built on the results of Dai *et al.* (2018) who ran an artefactual field experiment in public transport in Lyon using a similar subject pool as ours. Dai *et al*. (2018) estimated the proportion of dishonest individuals among fare-dodgers and non-fare-dodgers in a die-under-the-cup task. The estimated proportion of fully (partially) dishonest subjects was between 0% and 19% (41% and 60%) for non-fraudsters, and between 9% and 46% (37% and 74%) for fraudsters. Assuming similar proportions of full and partial liars in our field experiment, and assuming that full (partial) liars accept the banknote all (half of) the times, we predicted between 46% and 64.5% (30% and 39.5%) of fare-dodgers (non-fare-dodgers) taking the banknote. Using the midpoints of these intervals and assuming a type-I error rate of $\alpha = 0.05$ and a power level of 0.8, we computed a sample size of 92 subjects per group (fraudsters and non-fraudsters), which we rounded to 100 to be more conservative. Sample sizes are computed for two-sample proportions tests.

For the I condition, it was too speculative to make any prediction about the direction and the effect size from comparing NI to I. We thus set the sample size to 100 observations (*i.e.*, the optimal sample size for the NI condition) for each treated group (fraudsters and non-fraudsters) and computed the minimum detectable effect size for $\alpha = 0.05$ and power = 0.8. The minimum detectable effect size was 0.19 for fraudsters and 0.20 for non-fraudsters. This corresponds to a Cohen's *h* of approximately 0.4. Hence, a sample size of 100 was large enough to detect a small-medium treatment effect.

In running the field experiment, we thus decided to stop collecting data once we reached (roughly) 100 observations per group in the I condition. Collecting data in the I condition was much more complicated than in the other conditions since we were dependent on the natural occurrence of ticket inspections. Therefore, while we were trying to reach the target of 100 for the I condition, we continued to collect data in the NI condition (even if we had already collected 100 observations per group in this treatment) in order not to waste the actors' time (they were paid per hour). This is why we collected, overall, more data in the NI condition. The higher number of observations in the NI condition does not reflect any problem with the first hundred observations in this condition.

We excluded vulnerable persons, minors and tourists (based on subjective judgment), persons accompanied by children, friends, colleagues or partner. The research assistant collected background information about the subject pool through direct observation, as accurately as possible, during the field experiment. In Table A4 in Appendix 5 we provide detailed information about our subject pool. Subjects differ across conditions in terms of estimated age, ethnicity and wealth (cf. column 9). This is mainly due to the difference between fraudsters and non-fraudsters (cf. columns 10-11). Fraudsters tend to be younger (χ^2 test, $p = 0.006$ in I) and with a lower apparent wealth (χ^2 test, $p = 0.008$ in I). This is in line with previous evidence on public transport users (Dai *et al*., 2018). We also checked whether the samples differ between I and NI, and between NI and NI-A (*cf.* columns 12-15). The only statistically significant difference is in apparent wealth between I and NI for fraudsters ($p = 0.005$). Inspected fraudsters are perceived as poorer than non-inspected fraudsters. In all the other comparisons we cannot reject the null hypothesis that the data are independent across conditions. We can thus conclude that, except for the anticipated difference between fraudsters and non-fraudsters, the sample is fairly balanced across conditions. As a further check, we directly controlled for the individual characteristics of the subjects in the econometric analysis reported in Table 1.

APPENDIX 2: DETAILS OF THE LABORATORY EXPERIMENTS AND ADDITIONAL SURVEYS

2.1. LABORATORY EXPERIMENT 1

Laboratory experiment 1 was fully computerized using the z-Tree software (Fischbacher 2007), and conducted at GATE-Lab, Lyon (France). We recruited 45 subjects with the online software Hroot (Bock *et al.*, 2014). 46.67% of the subjects were males, 55.56% were students, and the average age was 28.18 (S.D. = 12.20).

Experimental Design. The experiment was divided into two parts. 15 subjects received only the second part. The remaining 30 subjects completed both parts. A translation of the instructions is provided in Appendix 3. On average, subjects who completed both parts earned €17.50 and those who completed only part 2 earned €14.07, including a €5 show-up fee.

In part 1, we employed the norm-elicitation procedure introduced by Krupka and Weber (2013). This elicitation method allows us to identify a social norm by measuring the collective perception regarding the appropriateness of a given behavior. Subjects were presented with two scenarios (one at the time) and asked to evaluate, on a four-point scale and for each scenario, whether the action taken by a person A was "very socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate" or "very socially appropriate". The incentives provided to the subjects were not to report their own preference but to match the response of the majority of subjects participating in the same session. The order of the two scenarios was randomized across subjects. In one scenario (henceforth Scenario 1), person A walks on the street, with no one around. She picks up a €5 banknote on the floor knowing that it does not belong to her. In the second scenario (henceforth Scenario 2), two strangers (A and B) walk on the street, with no one around. Person B picks up a €5 banknote on the floor and calls Person A's attention, asking whether she has lost it. Person A takes the banknote knowing that it does not belong to her. In both scenarios, subjects were asked to judge person A's decision to take the banknote. At the end of the experiment, one of the two scenarios was randomly drawn. For this scenario, if a subject's answer coincided with the answer given by the majority of all participants in the session, the subject earned €5.

In part 2, the task was to listen to 48 audio files chosen at random for each subject from all the recordings made with the phone by the actors during the field experiment. The recordings were randomly chosen such that, out of 48 audio files, we had 12 for each condition and group of subjects considered in our field experiment (excluding the No-Inspection-Audience condition). We explained to the subjects the context where the recordings were made but we did not tell them anything about the first stage of the field experiment. For each audio file, subjects listened to the voice of the actor asking whether the targeted person has lost the €5 banknote but not the answer of the person. They were asked to guess whether or not the targeted passenger took the banknote. Subjects were allowed to replay each audio file as many times as they wanted before reporting their guess. At the end of the experiment, we drew at random 5 guesses for each participant and paid him or her €4 for each correct guess.

Results of the Norm Elicitation Task. To test whether taking the banknote in our field experiment violates a social norm, we looked at how subjects judged person A's decision in Scenario 2. If person A's decision to take the banknote is perceived as unethical, people should rate it as inappropriate and they should value it as less socially appropriate than in the case where person A herself finds the banknote (Scenario 1). Following Krupka and Weber (2013), we assigned a score of –1 to "very socially inappropriate", –1/3 to "somewhat socially inappropriate", 1/3 to "somewhat socially appropriate", and 1 to "very socially appropriate". Table A1 in Appendix 5 presents, for each scenario, the average score and the frequency of each possible response. It reports the *p*-values of Wilcoxon signed-rank tests comparing the distributions of responses in the two scenarios, with each subject taken as an independent observation.

Table A1 confirms that taking the banknote in Scenario 2 (where the banknote is found by person B) is collectively considered as socially inappropriate. The mean score is negative and statistically different from 0 ($p < 0.001$). The modal response ("very socially inappropriate") receives 63% of the responses. No one judged the decision of person A in Scenario 2 as socially appropriate. We also observed that the decision of Person A in Scenario 2 is judged as less appropriate than in Scenario 1, and the difference is significant ($p < 0.001$). In Scenario 1, where the banknote is found by Person A, the decision to take the €5 is considered as "socially appropriate". The mean score is positive and statistically different from zero ($p < 0.001$), while the modal response is "somewhat socially appropriate", with 67% of subjects agreeing on that response. We can thus safely conclude that the decision to take the €5 banknote in our field experiment is collectively perceived as socially inappropriate and thus considered as a violation of a social norm.[20]

Results of the Guessing Task. The guessing task allows us to verify whether the actors played the scene similarly across conditions and groups, in which case the guessing ability of the subjects in the lab should be constant and not above chance levels. It can control for the actors' tone of voice and actual words spoken (but not for body language since it was forbidden to film the scenes). To determine whether the lab participants could predict the behavior of the subjects in the field, we constructed the following measure of the guessing ability (Belot and van de Ven, 2017):

$$A = F\,(T \mid T) - F\,(T \mid NT)$$

where $F\,(T \mid T)$ is the proportion of lab participants that guessed "took" when the field subject indeed took the banknote, while $F\,(T \mid NT)$ is the proportion of lab participants that guessed "took" when the field subject did not take the banknote. The advantage of this measure is that it is independent of the number of times the field subjects took the banknote. Depending on the value of A, we can make the following claims:

- If $A \leq 0$, the lab participants are not able to predict the behavior of the field subjects (*i.e.*, the probability of guessing that the person took the banknote is independent of the actual behavior of the person) or are worse than chance.
- If $0 < A < 1$, the lab participants can to some extent predict the behavior of the field subjects.
- If $A = 1$, the lab participants can perfectly predict the behavior of the field subjects.

Table A2 in Appendix 5 reports, for each condition and group of field subjects, (*i*) the proportions of lab subjects who guessed "took" when the field subjects took (first row) and did not take (second row) the banknote, respectively; (*ii*) the ability A to predict field behavior (third row); and (*iii*) the *p*-values of Wilcoxon signed-rank tests that $A = 0$ (fourth row). There is evidence that lab subjects were not able to predict the behavior of the field subjects. In all conditions and groups of field subjects, our measure of the ability to predict field behavior is either not significantly different from zero or (weakly) significantly negative (in the I condition for fraudsters), meaning that lab subjects were, if anything, worse than chance in predicting field behavior. We can also compare whether the ability to predict field behavior differs across conditions and groups of subjects. We find no evidence of any difference in pairwise comparisons (Wilcoxon signed-rank tests, $p > 0.1$ for all comparisons, taking the subjects' average A as the independent unit of observation).

Even if our lab subjects were not able to predict behavior in the field, they might still have perceived changes in the performance of the actors across conditions and groups of subjects, and assigned different probabilities of taking the banknote. To test this possibility, we ran a logit regression on the probability of guessing (correctly or not) that the person took the banknote. Independent variables are treatment dummies

[20] Similar results are obtained if we only consider the first scenario encountered by each participant and compare the scores of Scenarios 1 and 2 between subjects (Mann-Whitney test, $p < 0.001$). We can do that since each within-subject scenario was presented on a different computer screen and participants did not know about the content of the second scenario when they were responding to the first.

(using “NI fraudsters” as the baseline category) and fixed effects for the actors. Standard errors are clustered at the subject level. Table A3 in Appendix 5 shows the results of this regression. We find no evidence that subjects assigned a different probability of taking the banknote across conditions and groups of field subjects.

2.2. LABORATORY EXPERIMENT 2

The negative spillover effect of the deterrence institution that we observe in our field experiment may be the result of a change in the perception of the injunctive norm. Both fraudsters and non-fraudsters may, after an inspection, revise downward their beliefs about what ought to be done when the €5 banknote is proposed. To test this conjecture, we conducted Lab Experiment 2 with 96 participants from our subject-pool at GATE-Lab, Lyon (France). Subjects were recruited via the online software Hroot (Bock *et al.*, 2014). The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). Subjects were mostly students (92.71%), 56.25% were males, and the average age was 21.82 (S.D. = 6.71).

Experimental Design. There were two parts in the experiment. In part 1, subjects played a simplified version of the public transport game (Dai *et al.*, 2017). In this game, subjects had to make a risky choice which was described as the decision to buy or not a ticket for using a (fictional) bus, knowing that there was a risk of inspection. The ticket costed €1.8 (which was equivalent to the price of a ticket in Lyon when our field experiment was run). Each subject was inspected with 50% probability (this was randomly determined by the computer and it was independent for each subject). Subjects were not told about the precise probability of inspection (they only knew that there could be one). If a subject was inspected, the computer informed the subject about the inspection and displayed pictures and a video of real ticket inspectors in action to increase the salience of the event. An inspected subject who did not buy the ticket had to pay €4.80 (a fine of €3 plus the price of the unpaid ticket). There were no financial consequences for those who did not buy the ticket and were not inspected. Those who bought the ticket paid €1.8 both in the event of an inspection or no inspection. Any loss was deducted from the show-up fee which was purposely increased to €10 to make sure that subjects did not earn less than a minimal participation fee of €5.20.

In part 2, we employed the same norm-elicitation task that we used in Laboratory Experiment 1 (see above). Subjects earned on average €10.61. The instructions are provided in Appendix 3.

Results. We can test whether subjects who have been exposed to a ticket inspection changed their perception of the norm. For a clean test, we only focus on non-fraudsters (N = 56). Table A6 in Appendix 5 reports for each scenario, and distinguishing between non-fraudsters who were inspected in the game and those who were not, the average score of appropriateness and the frequency of each possible response in the Krupka-Weber task. It also reports the p-value of Wilcoxon rank-signed tests comparing the distributions of responses in the two scenarios, with each subject taken as an independent observation.

Table A6 in Appendix 5 confirms our results from Laboratory Experiment 1 that taking the banknote in Scenario 2 (B finds the banknote and asks A) is collectively considered as socially inappropriate while taking the banknote in Scenario 1 (A finds the banknote) is not. If we compare the behavior of inspected and non-inspected non-fraudsters, we do not find statistically significant differences. The mean score of appropriateness is similar between the two groups of subjects in both Scenario 1 (Mann-Whitney tests, p = 0.823) and Scenario 2 (p = 0.816). The distribution of responses is also not statistically different (Fisher's exact test, p = 0.784 and 0.744 for Scenarios 1 and 2 respectively). This suggests that, after an inspection, subjects do not revise their perception of the norm. The results are analogous if we only focus on the first scenario encountered by each participant.

2.3. SURVEY IN PUBLIC TRANSPORT

Design. To study whether inspections trigger negative emotions which may explain the negative spillover effect observed in our field experiment, we conducted a brief survey in the public transport of Lyon a few weeks after our field experiment. We instructed a research assistant to identify non-fraudsters travelling on board of buses and trams and approach them when they were getting off the public vehicle, using the same identification procedure as in our field experiment. The research assistant asked them whether they were willing to participate in a brief survey for researchers at the University of Lyon. The survey contained two simple questions using Self-Assessment Manikins (SAM) (Lang, 1980). The first question measured self-reported happiness on a scale from 1 (unhappy) to 5 (happy), while the second question measured emotional arousal on a scale from 1 (quiet, calm) to 5 (nervous). Each question was presented with five pictures associated to each possible answer, as shown in Figure A4 in Appendix 4. The passengers were instructed to answer each question using those pictures. The survey also elicited the beliefs of the passengers about the percentage of fare evaders in the transport network in Lyon. In particular, we asked interviewed people to indicate, in their opinion, how many passengers travel without a valid transport ticket/pass out of 100 passengers.

In total, the research assistant interviewed 109 non-fraudsters who had not been inspected and 51 non-fraudsters who had been inspected. Even if the focus was on non-fraudsters, she also collected data on 38 fraudsters who had been inspected.

Results. While we find a statistically significant difference both on the average level of happiness (2.26 and 3.88; Mann-Whitney test, $p < 0.001$) and nervousness (3 and 2.24, $p = 0.010$) between inspected fraudsters and inspected non-fraudsters, we find no difference between non-fraudsters who had been inspected and non-fraudsters who had not been inspected, both in reported happiness (3.88 and 4.14; $p = 0.257$) and nervousness (2.24 and 1.93; $p = 0.251$). Although we cannot exclude a selection bias in the willingness to take part in the survey, this suggests that inspections do not alter systematically the emotional state of individuals who validate their ticket in public transport.

If we look at the passengers' beliefs about the extent of fare evasion, we find that inspected fraudsters estimated a larger percentage of fare dodgers compared to both inspected (42.37% *vs*. 29.28%, $p = 0.003$) and non-inspected non-fraudsters (42.37% *vs*. 30.93%, $p = 0.001$). Instead, we find no difference in beliefs between inspected and non-inspected non-fraudsters ($p = 0.659$). This suggests that ticket inspection does not change the beliefs of non-fraudsters about the prevalence of fare evasion. All these results hold also if we look at the number of controllers conducting the ticket inspection (*i.e.,* the intensity of the inspection).

APPENDIX 3: INSTRUCTIONS OF LABORATORY EXPERIMENTS

3.1. INSTRUCTIONS OF LABORATORY EXPERIMENT

Instructions are translated in English from French.

Hello. Thank you for participating in this study. Please turn off your mobile phone. It is forbidden to communicate with other participants for the duration of the session.
If you have any questions at any time, please press the red button on the side of your desk and an assistant will come to answer your questions in private.
The experience is divided into two parts. At the end of the session, you will receive your earnings from parts 1 and 2 as well as a show-up fee of €5. Your earnings will be paid to you privately in a separate room to maintain confidentiality.
Please press OK to see the rest of the instructions.
{OK}

Part 1

Your task

The following screens will describe two situations in which a person "A" makes a choice. After you read the description of the situation, we will ask you to evaluate the choice made by person A and to indicate whether this choice is "socially appropriate" and "consistent with moral or proper social behavior in society" or "socially inappropriate" and "inconsistent with moral or proper social behavior in society". By "socially appropriate", we mean a behavior considered correct and ethical by the majority of people.
For each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behavior. To enter your response, you will have to click on one of the following options.

Person A's choice	**Very socially inappropriate**	**Somewhat socially inappropriate**	**Somewhat socially appropriate**	**Very socially appropriate**
	☐	☐	☐	☐

{OK}

Your earnings

At the end of the experiment, we will randomly select one of the two situations. For the situation selected, we will determine which response was selected by the largest number of participants in this session. If you give the same response as that most frequently given by the other participants, then you will receive an additional €5 which will be paid to you at the end of the session.
For instance, suppose that in the situation selected for the payment, your response had been "somewhat socially inappropriate", then you would receive €5 if this was the response selected by the largest number of participants in today's session.
If you have any questions, please press the red button on the side of your desk. Otherwise, press OK to start the task.

{OK}

Situation 1

Description: two persons (A and B) who do not know each other are walking on the street, with no one around. Person B walks behind person A. Person B picks up a €5 banknote from the ground and calls person A to ask if she has lost it. Person A takes the banknote knowing that it does not belong to her.
Please indicate whether you think that person A's choice is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate or very socially appropriate. To indicate your answer, click one of the options below. Remember that if this question is selected for the payment you will earn €5 if your response is the same as the most common response given by the other participants in today's session.

Person A's choice	Very Socially Inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
Take the banknote	☐	☐	☐	☐

{OK}

Situation 2

Description: One person (A) is walking on the street, with no one around. She picks up a €5 banknote from the ground knowing that it does not belong to her.
Please indicate whether you think that person A's choice is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate or very socially appropriate. To indicate your answer, click one of the options below. Remember that if this question is selected for the payment you will earn €5 if your response is the same as the most common response given by the other participants in today's session.

Person A's choice	Very Socially Inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
Take the banknote	☐	☐	☐	☐

{OK}

Part 2

Your task

Your task is to listen to 48 audio files. These audio files correspond to recordings made during a study conducted on the streets of Lyon in the following context.

For each recording, an actor or actress is at the phone. He/she is following a person on the street. After 20/30 meters, he/she catches up with the person and pretends to pick up a €5 banknote on the ground. The actor/actress calls the attention of the targeted person, holding for a moment his/her phone conversation, to ask if the person has lost the banknote. If the targeted person responds affirmatively, the actor/actress gives the banknote to the person. If the targeted person responds negatively, the actor/actress puts the banknote in his/her pocket. In both cases, the interaction between the person and the actor/actress stops, the actor/actress resumes his/her phone conversion and leaves.

Each recording was made by the actor/actress. The voice that you will hear is that of the actor/actress when he/she asks the targeted person if the €5 banknote belongs to him/her. The targeted person is not aware of the existence of the recording. You will not hear the answer of the targeted person. Some sentences may be different from each other but the context is always the same.

Your task is to guess, for each audio file, whether or not the targeted person took or not the €5 banknote.

You can replay each audio file multiple times before making your guess.
An example of the situation is accessible by clicking here:
{VIDEO}
{OK}

<u>Your earnings</u>

At the end of the session, the program will select at random five audio files. You will be paid for your guesses in these five audio files. For each selected audio file:

- you will earn €4 if your prediction is correct (i.e., you have correctly guessed whether the person took or not the €5 banknote);
- you will earn €0 if your guess is incorrect.

These earnings will be added to your other earnings of the session.
If you have any questions, please press the red button on the side of your desk. Otherwise, press OK to start the task.
{OK}

Audio file 1 of 48
Please click on "Listen" to play the audio file.
{Listen}
Your prediction:

O The person takes the €5 banknote
O The person does not take the €5 banknote

To what extent are you sure of your prediction, on a scale of 1 (totally uncertain) to 5 (totally certain)?
1 2 3 4 5

3.2. INSTRUCTIONS OF LABORATORY EXPERIMENT 2

Instructions are translated in English from French. We only report the instructions of Part 1. The instructions of Part 2 are similar to those used in Part 1 of Laboratory Experiment 1.

Hello. Thank you for participating in this study. Please turn off your mobile phone. It is forbidden to communicate with other participants for the duration of the session.
If you have any questions at any time, please press the red button on the side of your desk and an assistant will come to answer your questions in private.
The experience is divided into two parts. At the end of the session, you will receive your earnings from parts 1 and 2 as well as a show-up fee of €10. Your earnings will be paid to you privately in a separate room to maintain confidentiality.
Please press OK to see the rest of the instructions.

{OK}

Part 1

<u>Your task</u>

Imagine that you take a bus to reach a certain destination. Taking the bus requires you to buy a ticket that costs €1.8. There could be a ticket inspection on the bus. This inspection is determined by the computer program with a certain probability that you do not know. If you are not inspected or if you are inspected and you have bought a ticket, there is no consequence. If you are inspected and you have not bought a ticket, you will have to pay a fine of €3 and the price of the ticket (€4.8 in total).
Your task consists of deciding whether you want to buy the ticket or not. After your decision, there are four possible scenarios:
You have not bought the ticket and you are not inspected: your loss is €0.
You have not bought the ticket and you are inspected: your loss is €4.8.
You have bought the ticket and you are not inspected: your loss is €1.8.
You have bought the ticket and you are inspected: your loss is €1.8.
The losses of this part will be deducted from the show-up fee of €5.
Please click "OK" to make your decision.
{OK}

Decision

Click on "Ticket €1.8" if you want to buy the ticket or "No ticket" if you do not want to buy it.
You will know immediately if you are inspected. If you are not inspected, you will go directly to the next part.

{Ticket € 1.8}   {No ticket}

Inspection!

You are inspected!

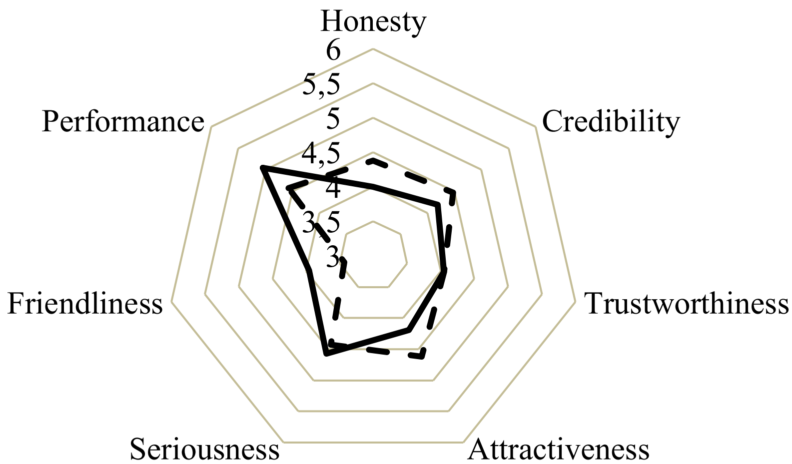
You have bought the ticket: there is no consequence [You have not bought the ticket: you pay a fine of €3 and the price of the ticket].
{OK}

In the on-screen original instructions, the picture in the middle is a video.

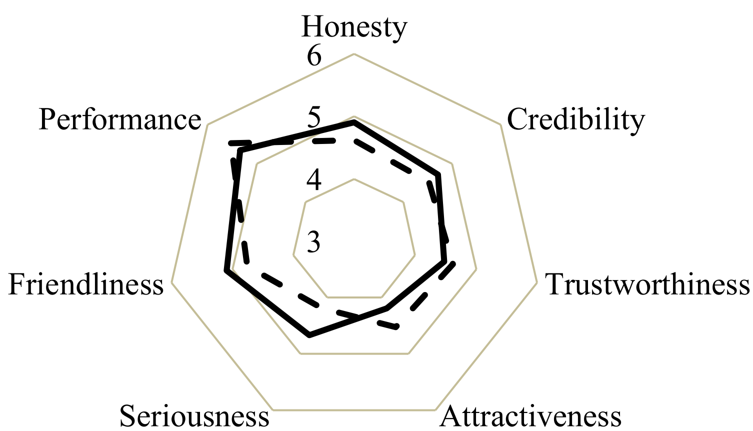
APPENDIX 4: FIGURES

Figure A1. Average scores of the four selected actors on the different characteristics.

Panel B: Lower-score actors

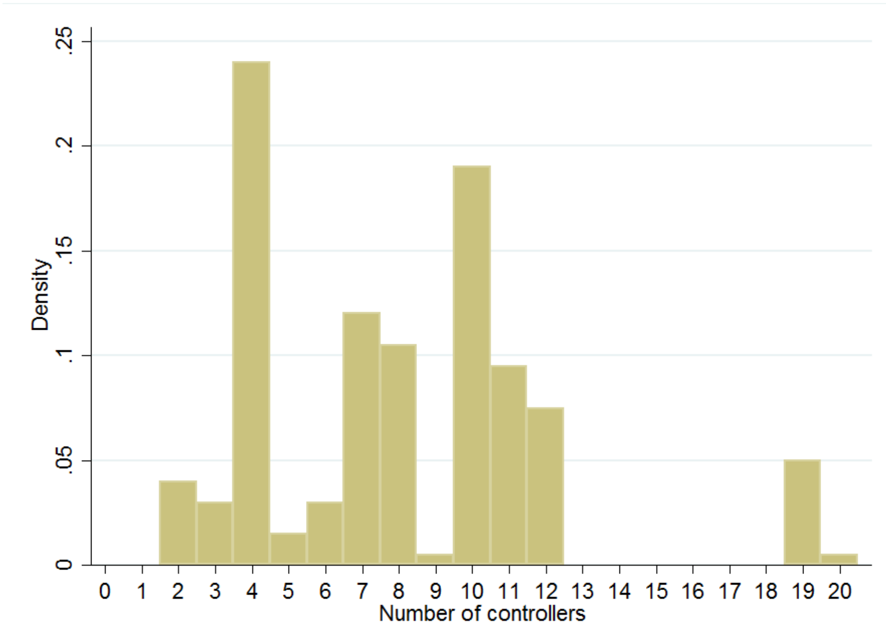


Panel A: High-score actors



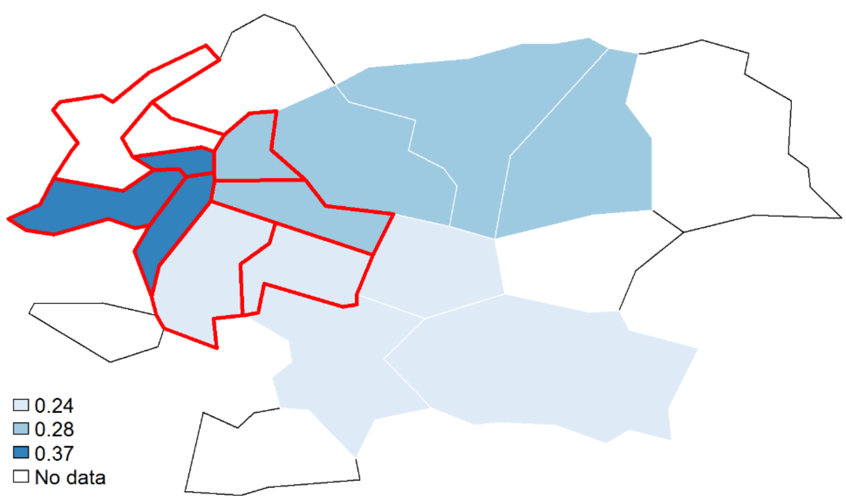
The figure depicts the average score given by student-subjects to each of the four selected actors. Panel A refers to the two actors with higher scores and Panel B refers to the two actors with lower scores. Each characteristic is measured on a scale from 1 to 7. The dashed line identifies an actress and the solid line an actor. This procedure ensured a neutral selection of the actors. In the main econometric analysis we introduced individual fixed effects to control for the different observable and unobservable characteristics of the actors.

Figure A2. Histogram of the number of ticket inspectors in the field experiment.



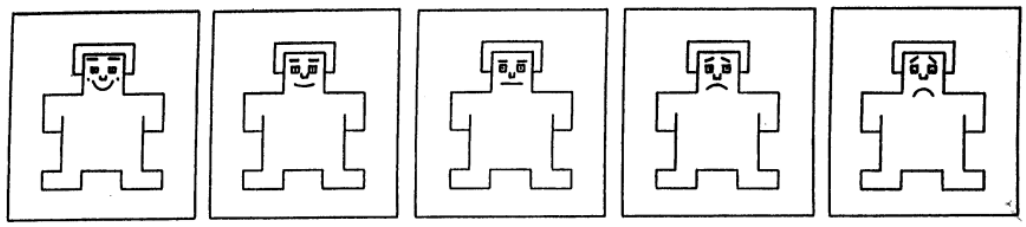
The figure shows the histogram of the number of ticket inspectors (per inspection) that we encountered in the field experiment (N = 200).

Figure A3. Frequency of ticket inspections in the bus-tram area of Metropolitan Lyon, as measured in our field experiment.

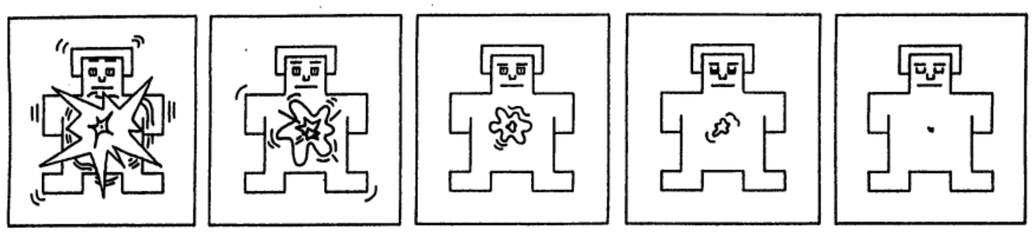


The figure plots the frequency of observed ticket inspections onto a map covering the area of Metropolitan Lyon that can be reached by tram or bus. The darkest blue shaded area identifies Center Metropolitan Lyon; the medium blue shaded area corresponds to Nord-East Metropolitan Lyon; while the light blue shaded area represents South-East Metropolitan Lyon. White segments are areas which we did not cover for logistical reasons. The area with red contours identifies the city of Lyon.

Figure A4. Self-Assessment Manikin (SAM) questions.



5	4	3	2	1



5	4	3	2	1

The figure depicts the pictures used in the survey to measure the emotional state (happiness in panel A and nervousness in panel B) of the participants on a scale from 1 to 5.

APPENDIX 5: TABLES

Table A1. Appropriateness scores across scenarios (N = 30) in Laboratory Experiment 1.

The Table reports the mean responses to the task eliciting the appropriateness of behavior in two scenarios. "Very socially inappropriate" (– –); "somewhat socially inappropriate" (–), "somewhat socially appropriate" (+), "very socially appropriate" (++).To construct the mean score, we assign a value of –1 to "very socially inappropriate", –1/3 to "somewhat socially inappropriate", 1/3 to "somewhat socially appropriate" and 1 to "very socially appropriate". Modal responses are shaded in grey.

Scenario	Mean	$--$	$-$	$+$	$++$	Rank-sum test
B finds the banknote and asks A	−0.76	63.33%	36.67%	0%	0%	$p < 0.001$
A finds the banknote	0.29	0%	20%	66.67%	13.33%	

Table A2. Proportion of lab subjects who guessed "took" depending on the actual behavior of the field subjects (N = 45) in Laboratory Experiment 1.

The Table reports the ability to predict field behavior, as measured by A=F(T | T)-F(T|NT), with F (T | T) the proportion of lab participants that guessed "took" when the field subject took the banknote and F (T | NT) the proportion of lab participants that guessed "took" when the field subject did not take the banknote. Standard errors are reported in parentheses. The last row reports the *p*-values of two-tailed Wilcoxon signed-rank tests using the mean prediction of each lab subject as the independent unit of observation. The null hypothesis is A = 0.

Conditions and groups	All	NI fraudsters	NI non-fraudsters	I fraudsters	I non-fraudsters
Subject took, $F(T\|T)$	0.51 (0.02)	0.48 (0.03)	0.53 (0.04)	0.54 (0.02)	0.49 (0.03)
Subject did not take, $F(T\|NT)$	0.51 (0.02)	0.49 (0.03)	0.5 (0.03)	0.6 (0.04)	0.51 (0.03)
Ability to predict (A)	0.00	−0.01	0.03	−0.06	−0.02
Wilcoxon test, $A = 0$ (p-value)	0.969	0.852	0.955	0.053	0.663

Table A3. Effect of main treatments on guessing that a person took the banknote in Laboratory Experiment 1.

The Table reports the average marginal effects of logit estimates. Standard errors are clustered at the individual level to account for the fact that subjects made 48 guesses. The guess that the person in the field took the banknote is regressed on treatment dummies (using NI fraudsters as a baseline). We included a dummy for each actor. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (Wald tests).

Dependent variable: 1 if subject guessed that the person took the banknote	dy/dx	Std. Err.	p>z
NI, non-fraudster	0.00	0.03	0.925
I, fraudster	0.05	0.03	0.129
I, non-fraudster	0.01	0.03	0.653
Actor fixed effects	Yes		
Number of observations	2160		
Pseudo R^2	0.006		
Prob > chi^2	0.088		

Table A4. Descriptive statistics on targeted passengers' individual background variables in the field experiment.

The Table presents the mean individual characteristics of targeted passengers in the field experiment by group and by condition. Columns (8) to (14) present the P-values for the null hypothesis that the data for each variable are independent across conditions (two-sided χ^2 tests). In column (8), the tests are conducted across all conditions. In columns (9)-(14), the tests are based on pairwise comparisons. N = 708.

	NI		**I**		**NI-A**			**χ^2 test (*P*-value)**						
Characteristics	**Fraudsters**	**Non-fraudsters**	**Fraudsters**	**Non-fraudsters**	**Fraudsters**	**Non-fraudsters**	**Total**	**All**	**NI fraud. *vs.* NI non-fraud.**	**I fraud. *vs.* I non-fraud.**	**NI fraud. *vs.* I fraud.**	**NI non-fraud. *vs.* I non-fraud.**	**NI fraud. *vs.* NI-A fraud.**	**NI non-fraud. *vs.* NI-A non-fraud.**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Gender														
Female	44%	51%	47%	44%	45%	52%	47%	0.758	0.281	0.691	0.677	0.316	0.854	0.869
Male	56%	49%	53%	56%	55%	48%	53%							
Age														
18-24	29%	23%	38%	26%	35%	24%	29%	**0.027**	0.527	**0.006**	0.400	0.320	0.606	0.529
25-34	33%	29%	31%	20%	26%	26%	28%							
35-44	12%	17%	9%	13%	16%	19%	15%							
45-59	16%	19%	17%	22%	15%	24%	19%							
≥ 60	10%	13%	5%	19%	8%	7%	10%							
Ethnicity														
Caucasian	54%	69%	55%	63%	52%	65%	60%	**0.061**	0.117	0.447	0.536	0.344	0.119	0.540
Arab	19%	13%	24%	16%	21%	12%	17%							
African	19%	12%	16%	18%	23%	19%	18%							
Asian	6%	4%	2%	1%	0%	3%	3%							
Other	3%	2%	3%	1%	5%	1%	2%							
Wealth														
Poor	21%	15%	40%	21%	25%	19%	23%	**0.002**	0.325	**0.008**	**0.005**	0.410	0.482	0.475
Average	71%	74%	54%	66%	71%	73%	69%							

Wealthy	8%	11%	6%	13%	5%	8%	8%							
Religious signs														
No	97%	98%	96%	98%	98%	98%	98%	0.886	0.702	0.380	0.627	0.905	0.594	0.825
Yes	3%	2%	4%	2%	2%	2%	2%							

Table A5. Determinants of the decision to take the banknote (controlling for the presence of an experimenter).

Dependent variable: Decision to take the banknote	Model (2)		Model (3)		Model (4)	
	dy/dx	S.E.	dy/dx	S.E.	dy/dx	S.E.
Treatment (baseline = Fraudster NI)						
Non−fraudster NI	−0.19***	0.06	−0.18***	0.06	-	-
Fraudster I	0.14**	0.07	0.14**	0.07	-	-
Non−fraudster I	−0.03	0.07	−0.04	0.07	-	-
Fraudster NI-A	−0.05	0.07	−0.06	0.07	-	-
Non−fraudster NI-A	−0.13**	0.07	−0.16**	0.07	-	-
Actors/Actress (baseline = Higher−score actress)						
Lower−score actress	0.20***	0.05	-	-	0.19***	0.05
Higher−score actor	−0.03	0.07	-	-	−0.03	0.07
Lower−score actor	0.09*	0.05	-	-	0.08	0.05
Male passenger	0.02	0.04	-	-	0.02	0.04
Age of the passenger [e]	0.03**	0.01	0.03*	0.01	0.03*	0.01
Ethnicity of the passenger *(baseline = Caucasian)* [e]						
Arab	0.06	0.05	0.03	0.05	0.05	0.05
African	0.06	0.05	0.05	0.05	0.05	0.05
Asian	0.08	0.11	0.08	0.11	0.05	0.11
Other	0.09	0.11	0.12	0.12	0.08	0.11
Social appearance of the passenger (baseline = poor) [e]						
Average	−0.20***	0.05	−0.19***	0.05	−0.19***	0.05
Wealthy	−0.25***	0.08	−0.25***	0.08	−0.25***	0.08
Religious signs *(baseline = no signs)*	−0.21**	0.1	−0.20**	0.1	−0.19*	0.1
Bus *(baseline = Tram)*	0.03	0.08	0.03	0.08	0.05	0.08
Many people in the vehicle *(baseline = few people)*	−0.05	0.04	−0.07*	0.04	−0.05	0.04
Weather (baseline = sunny)						
Cloudy	−0.02	0.05	−0.05	0.05	−0.02	0.05
Rainy	−0.08	0.09	−0.14	0.09	−0.09	0.09
Someone could notice the scene *(baseline = no one)*	−0.04	0.04	−0.04	0.04	−0.03	0.04
Center Metropolitan Lyon	0.06	0.04	0.06	0.04	0.06	0.04
Main Line	0	0.06	0.03	0.06	−0.04	0.06
Afternoon (baseline = morning)	−0.07*	0.04	−0.05	0.04	−0.07*	0.04
Gender interaction (baseline = Female actress, Female passenger*)*						
Female actress, Male passenger	-	-	0.03	0.05	-	-
Male actor, Female passenger	-	-	−0.05	0.06	-	-
Male actor, Male passenger	-	-	−0.02	0.05	-	-
Non−fraudster *(baseline = fraudster)*	-	-	-	-	−0.16***	0.04
Number of ticket inspectors	-	-	-	-	0.02***	0.01
Audience *(baseline = no audience)*	-	-	-	-	0	0.05
Presence of the experimenter	0	0.05	0.03	0.05	−0.03	0.05
Number of observations	708		708		704[§]	
Prob > chi2	<0.001		<0.001		<0.001	
Pseudo R2	0.111		0.090		0.110	

Notes: The Table reports the average marginal effects from Logit estimates. The dependent variable is the decision to take the banknote (= 1 if the banknote is taken and 0 otherwise). [e] Estimated by the research assistant. § Four observations were dropped because information about the number of inspectors was missing. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (Wald tests).

Table A6. Appropriateness scores across scenarios and conditions in Laboratory Experiment 2.

The Table reports the mean responses to the task eliciting the appropriateness of behavior in two scenarios. "Very socially inappropriate" (– –); "somewhat socially inappropriate" (–), "somewhat socially appropriate" (+), "very socially appropriate" (++). To construct the mean score, we assign a value of –1 to "very socially inappropriate", –1/3 to "somewhat socially inappropriate", 1/3 to "somewhat socially appropriate" and 1 to "very socially appropriate". Modal responses are shaded in grey. N = 96.

Situation and scenario	Mean	− −	−	+	+ +	Rank-sum test
Non-fraudsters no inspection (31)						
B finds the banknote and asks A	−0.57	41.94%	54.84%	3%	0%	$p < 0.001$
A finds the banknote	0.53	0%	6.45%	58.06%	35.48%	
Non-fraudsters inspection (25)						
B finds the banknote and asks A	−0.57	48%	44%	4%	4%	$p < 0.001$
A finds the banknote	0.49	0%	12%	52%	36%	