

FRAUD DETERRENCE INSTITUTIONS REDUCE INTRINSIC HONESTY*

Fabio Galeotti, Valeria Maggian and Marie Claire Villeval

The impact of deterrence institutions beyond their immediate scope of application is unknown. Using a quasi-experiment with naturally occurring variation in inspections, we found evidence of spillover effects across contexts. We identified fraudsters and non-fraudsters on public transport who were or not exposed to ticket inspections. We measured the intrinsic honesty of the same persons in a new context where they could misappropriate money. Instead of having an educative effect across contexts, the exposure to deterrence practices increased unethical behaviour of fraudsters but also, strikingly, of non-fraudsters. Learning about the prevailing norm is the likely channel of this spillover effect.

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 internalisation mechanism that induces individuals to comply even in the absence of any threat of punishment (Axelrod, 1986; Gintis, 2003). However, intrinsic honesty is not sufficient to prevent violations, and varies widely across cultures (Fisman and Miguel, 2007; Gächter and Schulz, 2016; Cohn *et al.*, 2019). While peer punishment of violations facilitates compliance (Fehr and Gächter, 2000; Masclet *et al.*, 2003), inspections and sanctions by centralised authorities are the most common institutional practices adopted in modern societies to deter deviant behaviour. When they raise the costs of breaking the rule above its benefits, these institutions can discourage the targeted misbehaviour (Becker, 1968; Di Tella and Schargrodsky, 2004; Fisman and Miguel, 2007; Baldassarri and Grossman, 2011; Ariely, 2012; Chalfin and McCrary, 2017). However, they sometimes crowd out the intrinsic motivation to comply (Frey, 1997; Gneezy and Rustichini, 2000; Frey and Jegen, 2001; Falk and Kosfeld, 2006; Dickinson and Villeval, 2008) with potential spillovers into adjacent activities (Belot and Schröder, 2016). This results from control-averse

* Corresponding author: Marie Claire Villeval, Univ Lyon, CNRS, GATE, UMR 5824, 93, Chemin des Mouilles F-69130 Ecully, France, and IZA, 53113 Bonn, Germany. Email: villeval@gate.cnrs.fr

This paper was received on 16 September 2019 and accepted on 27 February 2021. The Editor was Rachel Kranton.

The data and codes for this paper are available on the Journal website. They were checked for their ability to reproduce the results presented in the paper.

We thank J. Andreoni, M. Bigoni, G. Charness, 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), from the INDEPTH program of IDEXLYON (No 183634), and from a grant of the French National Research Agency (ANR, DECISION project, ANR-19-CE26-0019). This project has also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 661645. The IRB of INSERM (IRB00003888) has approved the project (decision n°17-416). The AEARCT identification number of the present study is AEARCTR-0004656.

individuals who directly reciprocate against a distrustful authority that reduces their freedom of choice.

While past estimations of deterrence effects focus almost exclusively on the targeted misbehaviour, we contend 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 prosocial 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 behaviour beyond their immediate scope of intervention. Yet, 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 behavioural economic theories could account for these effects—which may be positive or negative. Focusing on dishonest individuals, if past experience of a deterrence institution recalls what society expects from individuals, it may serve 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.*, 2014). 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 cleansed 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, signalling theories (e.g., Benabou and Tirole, 2003) deliver mixed predictions: an inspection may signal to compliers that they are honest, reinforcing their intrinsic motivation; but it may also remind some people that their motivation for compliance is avoiding a fine, and crowds out their honesty in subsequent contexts where they know the deterrence institution is not in place anymore. Also, because of social learning, the enforcement of the deterrence institution may affect compliers' beliefs about the spread of norm violation

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

² Most of the literature on deterrence in the Beckerian tradition examines whether variations in the probability of detection versus severity of sanctions affect criminality (see Chalfrin and McCrary, 2017 for a survey). The only spillovers considered are those related to crime displacement following a sudden increase in the intensity of policing (see the 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 consider only 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.

in society. Observing many violators being punished or large police teams may reveal that misconduct is socially widespread and has become the norm, which may lead compliers to behave accordingly (Keizer *et al.*, 2008; Dickinson *et al.*, 2015).

In sum, there exist 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 affect rule compliers and non-compliers alike remain open questions. To shed light on these speculations, we ran a quasi-experiment in public transport with naturally occurring variation in inspections and on the streets in Lyon, France, with 708 passengers. We collected a direct and unbiased measure of dishonest behaviour (i.e., fare evasion). Individuals were observed in a daily-life situation and were not aware of their taking part in a study. Our quasi-experiment allows us to overcome the limitations associated with laboratory experiments (Levitt and List, 2007; List, 2011) when investigating dishonesty, since possible scrutiny by the experimenter can considerably influence unethical behaviour (Gneezy *et al.*, 2018). Finally, besides contributing to the analysis of the dynamics of unethical behaviour (but in a different sense to Welsh, 2015, and Garrett *et al.*, 2016, who looked at escalation effects), we focus on both compliers and non-compliers.

We chose to conduct our study 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. Moreover, in this setting dishonest behaviour is publicly identifiable with almost no measurement error, since every passenger must validate a ticket or a pass every time they enter a public vehicle.

Our quasi-experiment consisted of two stages. The first stage took place on board buses and trams and produced two main natural conditions. In the *inspection condition*, an inspection by ticket inspectors from the transport company naturally occurred during the journey whereas in the *no-inspection condition*, no ticket inspection occurred.⁴ The second stage took place when the passenger exited the vehicle, on the street. A professional actor who was part of the research team walked behind the targeted passenger and suddenly bent down to seemingly pick up a banknote on the ground. The actor then called the passenger's attention by asking whether they had lost the banknote. We measured intrinsic honesty by recording whether or not the passenger took the banknote, and tested whether this correlates with their compliance on public transport. To identify the causal effect of the deterrence institution on compliers and non-compliers, we contrasted intrinsic honesty in the inspection versus no-inspection conditions. We explored the possible mechanisms behind our findings by means of an additional survey conducted in public transport, and two laboratory experiments.

We found that instead of having a positive immediate educative effect in the new context, the direct exposure to a deterrence institution in public transport increased the misconduct of fraudsters on the street. More strikingly, it also significantly increased the unethical behaviour of

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 (http://www.sytral.fr/uploads/Externe/9d/310_765_CP_CS_02_02_2018.pdf, accessed 23 September 2020).

⁴ Note that our identification strategy is not based on exogenous shocks in the deterrence policy but on natural variations in its implementation. 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. Since inspections result from the transport company policy, and not from a random intervention by the researchers, our study can be defined as a quasi-experiment. We address the potential issues regarding randomisation in Section 1.

non-fraudsters. The effect was highly significant and was of the same magnitude for both groups (between 14% and 19% of the base level). This rejects a general explanation of the spillover in terms of monetary loss recovery. Interestingly, the effect increased with the size of the ticket inspection team, especially for non-fraudsters. Without rejecting the role of emotions in affecting our results, this suggests that one mechanism behind such an effect may be a normative channel: larger inspection teams may signal more widespread dishonesty and a weaker social norm of honesty in the society. Overall, our findings show that to optimise the design of deterrence institutions and evaluate their full efficiency, policymakers should also consider the spillover effects of these institutions on intrinsic honesty beyond the context where these institutions directly apply.

1. Experimental Design

Our quasi-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 naturally enforced. It took place on public transport in 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 fareboxes located on board public vehicles each time they enter a new vehicle. Re-validation of the same ticket is compulsory even if the ticket is valid for one hour from the first validation and has already been 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 Details about the public transport network in Lyon and ticket inspections can be found in Online Appendix A.1.

1.1. First Stage

In the first stage of the study, the research assistant and the actor travelled on board a bus or tram. The former had to stay next to a validating farebox and focus attention on the first four of five passengers boarding and validating or not validating their ticket. This was done for logistical reasons: first, it was easier to recognise those who validated and those who did not by targeting the first passengers entering the public vehicle; second, it was easier to recall who stamped the ticket and who did not if only a few passengers at a time were targeted. The actor waited on board

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 recorded casting with eighteen candidates in a professional acting school. Twenty-one 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, apparent honesty, trustworthiness, attractiveness, credibility, seriousness and friendliness, based on the videos. We selected those actors with similar high scores in performance and credibility (see Figure D1 in Online Appendix D). Before us, only a 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 31% of the observations were 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.

⁷ After scanning a ticket or a pass, the farebox 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 for 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.

the public vehicle without giving any 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. This procedure avoided subjectivity in the choice of the target passenger and was cost effective, preventing the research team from spending too much time on board.

There were two conditions that occurred naturally. In the inspection condition (I condition, 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 condition, hereafter), no ticket inspection naturally occurred. We address the question of randomisation in a separate subsection below.

1.2. *Second Stage*

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 minimise 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 they had lost the banknote. Accepting or declining 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 decline 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 within hearing distance. As a robustness check of the effect of an audience on intrinsic honesty we ran an additional condition, the no-inspection-audience condition (NI-A condition, hereafter). Here, the assistant walked by the actor and explicitly observed the scene. This allowed us to isolate the possible role played by an observer in influencing individuals' unethical behaviour.

The actors were asked to use their mobile phone as an audio recording device when playing the scene on the street.⁸ In order to make sure that (i) the actors played the scene similarly across conditions, and (ii) one's false claim of ownership of a banknote found by someone else violates an injunctive ethical norm, we conducted a laboratory experiment (called 'laboratory experiment 1' that is presented in detail in Online Appendix B.1 (see the instructions in Online Appendix C.1)).

1.3. *Recorded Variables*

The research assistant had to record several pieces of information on a tablet. He recorded the name of the actor playing the scene, the time of day, the weather (sunny, cloudy or rainy), the bus/tram line, whether the subject validated a ticket or a monthly pass or nothing, the bus/tram stop where the subject got off the public vehicle, the approximate number of people on board (almost empty, quite crowded but everyone could sit, crowded), whether someone could notice the scene played in the street, whether or not the subject took the banknote, the gender, estimated

⁸ These recordings were used to verify that the actors played the scene according to the protocol (see below), and as a robustness check to ensure that any minimal deviation from this protocol did not affect the internal validity of our results. We thank James Andreoni for suggesting this to us. The script given to the actors is available in Online Appendix A.2.

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. 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 plain clothes, the gender of the controller who inspected the targeted passenger and, only for inspected fraudsters, whether the passenger paid the fine immediately, and whether he or she had an emotional or aggressive reaction to being fined.

Tables E1 and E2 in Online Appendix E present descriptive statistics of the targeted passengers' individual background variables in each condition for the whole sample, and for the sub-groups of fraudsters and non-fraudsters, respectively.

1.4. *Sample Size and Power Analysis*

The experiment was run on weekdays in 2017. On a typical weekday, we collected on average twenty observations between 9:00 a.m. and 6:15 p.m., avoiding rush hours because passengers may anticipate that the risk of ticket inspection is lower during these hours. In total, our study involved 708 passengers: 358 non-fraudsters (104 in the I condition, 140 in the NI condition and 114 in the NI-A condition) and 350 fraudsters (100 in the I condition, 140 in the NI condition, and 110 in the NI-A condition). When collecting data, we excluded vulnerable persons, minors and tourists (based on subjective judgement), persons accompanied by children, friends, colleagues or partners. Details about the sample distribution across lines and at different times of the day and locations are reported in Table E3 in Online Appendix E, and Figure D4 in Online Appendix D displays the frequency of inspections on a map.

To determine the sample size for both the NI and I conditions, we conducted an a priori power analysis. For the NI condition, we built on the results of Dai *et al.* (2018) to achieve a sample size of ninety-two subjects per group (fraudsters and non-fraudsters), which we rounded to a more conservative one hundred (see details in Online Appendix A.3). For the I condition, it was too speculative to allow any prediction about the direction and the effect size by comparing the NI condition to the I condition. We thus set the sample size at one hundred 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 corresponded to a Cohen's *h* of approximately 0.4. Hence, a sample size of one hundred was large enough to detect a small-medium treatment effect. In running the quasi-experiment, we thus decided to stop collecting data once we reached (roughly) one hundred observations per group in the I condition.

Collecting data in both conditions was time consuming, especially in the I condition, which depended on the natural occurrence of ticket inspections.⁹ Therefore, we instructed the research assistant and the actor to primarily focus on searching for ticket inspectors by travelling up and down a random line and switching to another if unsuccessful. They were asked to start from a

⁹ By choosing a target of one hundred in each condition, we obviously over-represented the population of inspected passengers. In fact, the probability of being inspected is quite low in the field. Egu and Bonnel (2020) estimated that in 2017 in Lyon, the ratio between the number of ticket inspections and the number of boardings was 0.017 for the tram and 0.012 for the bus. Precisely, boardings amounted to 95.2M and 166.1M for the tram and bus networks, respectively, as measured by the counting system placed at each vehicle door; controllers inspected 1.6M and 1.9M people in the tram and bus, respectively. Thus, respecting this proportion with the constraint of one hundred observations in the I condition would have required collecting between about 5,882 and 8,333 observations in the NI condition.

different main line every day in a direction chosen at random, with the overall objective of keeping a roughly constant proportion of I and NI observations throughout the day. While endeavouring to reach the target of one hundred for the I condition, we collected, however, more data in the NI condition so as not to waste the actors' time (they were paid per hour). It is important to note that every day we sampled roughly three or four random observations in the NI condition for each observation sampled in the I condition, to account for the different sampling costs, and tried to maintain this ratio throughout the experiment.¹⁰ The higher number of observations in the NI condition does not reflect any problem with the first hundred observations in this condition.

1.5. Identification

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 two-fold 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 inspections on the latter. Our identifying strategy relies on the assumptions that our sampling of participants is random and that ticket inspections are orthogonal to intrinsic honesty.

Regarding the first assumption and the first dimension (fraudsters versus non-fraudsters), there is no reason to believe that the order in which passengers board a public vehicle correlates with their intrinsic honesty. By focusing on the first four or five passengers boarding we could randomly observe fraudsters or non-fraudsters. But one could question whether the rule pertaining to targeting the first pre-identified passenger getting off the vehicle generated a lack of generalisability by focusing on short trips. In fact, we do not necessarily over-represent short trips. According to Egu and Bonnel (2020), passengers in Lyon change between 1.30 and 1.45 vehicles per trip. So, when a passenger enters a new vehicle, it may be the final leg of a longer trip.¹¹ Moreover, while the motivation to defraud may be different if one considers a long, rather than short, trip, it is not clear how this would affect people's reaction to a ticket inspection in terms of acceptance of the banknote and how long a trip need be to observe a different response.

Regarding the other dimension (inspected versus non-inspected passengers), imposing a ratio between observations in the I and NI conditions in the data collection, as explained above, circumvented collecting all the NI observations immediately, and the I observations later, which could have raised selection issues. Moreover, to verify that our randomisation strategy worked, we checked with the transport company the consistency between the frequency of inspections observed in our data and those reflected in the inspection plans of the company. We found that the distribution of the inspections over time has an inverted U-shape pattern analogous to that observed in our field data.¹² A regression analysis of the occurrence of a planned inspection by the company at the time of our experiment provided results highly consistent with our data, suggesting that there was no bias in the method of our collecting the I condition observations.

¹⁰ Imposing this 3:1 or 4:1 ratio allowed us to account for the substantially different costs of sampling in the two conditions. Indeed, it has been shown that the optimal sample sizes should be inversely proportional to the square root of the relative sampling costs (see Pentico, 1981; List *et al.*, 2011). Since the sampling costs are much higher in the I condition, the NI sample should be larger than the I sample.

¹¹ Because the bus trip may follow another trip with the same ticket (since a ticket is valid for one hour), it is not the case that the experiment focuses on people who pay a ticket for a short trip.

¹² The inspection plans of the company cannot be made publicly available, but this analysis was provided to the referees.

The second identification assumption (orthogonality of ticket inspections to intrinsic honesty) would automatically hold if ticket inspections were completely random across lines and times of day. However, the randomisation of inspections is not perfect since they result from the company policy. While this, in itself, does not pose a threat to our identification strategy as long as ticket inspections are not systematically conducted in areas where or at times when intrinsic honesty is particularly low, it might be a source of concern were there an asymmetric selection in the samples of inspected and non-inspected people. A first important point is that ticket inspections are organised by the company such that they are difficult to predict by passengers, a major source of randomness. As explained by the company, inspections are largely random in order to maintain uncertainty and prevent fare dodgers learning where and when inspections could happen. Some randomness also results from the fact that if inspection teams receive an inspection plan at the beginning of their shift, they often change that plan to adapt to the occurrence of incidents. Indeed, they are also in charge of the security of the system: in the case of an incident on a line, they may switch tasks and lines independently of the initial plan.

Table E3 in Online Appendix E shows that ticket inspections were more frequent in the Centre Metropolitan Lyon, on certain lines and in the early afternoon. Therefore, we include, in the parametric models reported in the results section, fixed-effect variables (time of day, geolocation and transport line) that might be a source of selection for inspections, as detailed in the next section. To assess the robustness of our results, a series of additional steps is then implemented. In Tables E8 and E9 of Online Appendix E, we consider additional regression models where we include finer control that takes into account the environmental variables as well as interactions between each transport line and the time of day, and interactions between the geolocation and the time of day. In Online Appendix A.4 we also report a propensity score matching analysis to account for possible selection on observables. Indeed, if ticket inspectors act in a non-random way, their selection criteria should be mainly based on observable characteristics of a transport (a specific area, line or time), which limits the risk of selection on unobservables. They are unlikely to base selection on observable individual characteristics, since all passengers are checked in the case of an inspection. For example, we tested for correlation between the size of the inspection teams and the individual characteristics of the subjects in terms of gender, age and ethnicity and found none (available upon request). The results of our robustness checks confirm our main analysis, suggesting that differences in timing and location of the controls are unlikely to be responsible for our results.

2. Results

Our first result shows that not validating the ticket on 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 displays the percentage of fraudsters and non-fraudsters who took the banknote, depending on whether a ticket inspection occurred (I) or not (NI). In the NI condition, 52.86% of the fraudsters took the banknote compared to 32.14% of the non-fraudsters. This difference is significant (χ^2 -squared test: $\chi^2(1) = 12.29$, $p < 0.001$),¹³ revealing that the disparity in ethical behaviour correlates

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

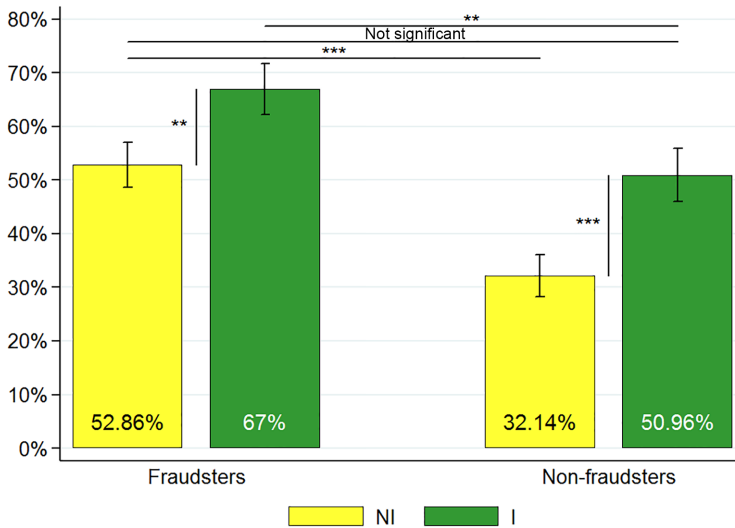


Fig. 1. Percentage of Fraudsters and Non-fraudsters Accepting the Banknote in the Conditions with (I) and without (NI) Ticket Inspection.

Notes: The light bars are for the NI condition and the dark bars for the I condition. Here $N = 140$ (NI, fraudsters), 100 (I, fraudsters), 140 (NI, non-fraudsters) and 104 (I, non-fraudsters). Error bars, mean \pm SEM. Significance levels: *** $p < 0.01$, ** $p < 0.05$, χ^2 -squared tests.

across the two contexts. The observed pattern of cross-context unethical behaviour was not affected by the presence of an observer in the second stage (see Figure D2 in Online Appendix D). Although fraudsters were slightly less likely to take the banknote when being observed by a third person (45.45% in the NI-A condition versus 52.86% in the NI condition), the difference is not significant ($\chi^2(1) = 1.35$, $p = 0.245$). Similarly, there was little difference in the percentage of non-fraudsters who took the banknote in the NI-A and NI conditions (33.33% versus 32.14% in the NI condition; $\chi^2(1) = 0.04$, $p = 0.841$).

Next, we investigated the effect of the enforcement of the deterrence institution in the first stage of the quasi-experiment on the intrinsic honesty of fraudsters in the second stage (i.e., behaviour on the street). This spillover effect was 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 had no immediate educative effect on the intrinsic honesty of fraudsters.

Fraudsters caught travelling irregularly had to pay a fine. Hence, the mechanism behind this negative spillover could be that fraudsters try to partially recover the loss incurred by the fine (Sharma *et al.*, 2014). However, if loss recovery solely explained 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 a ticket inspection ($\chi^2(1) = 8.79$, $p = 0.003$). Thus, the enforcement of the deterrence institution also reduced the intrinsic honesty of law-abiding passengers. The percentage of passengers who took the banknote following an inspection was still significantly higher for fraudsters than for non-fraudsters ($\chi^2(1) = 5.41$, $p = 0.020$), but between non-fraudsters after an inspection and non-inspected fraudsters it was no longer different, even without an audience ($\chi^2(1) = 0.09$, $p = 0.769$).

We now turn to a regression analysis to control for the environment and for the individuals' socio-demographic characteristics. The coefficients from four linear probability regressions in which the dependent binary variable is the decision about whether or not to take the banknote are reported in Table 1.¹⁴ In model (1), the effect of inspections is investigated: the independent variable *inspection* is equal to 1 when a control occurred on the bus/tram and is equal to 0 otherwise; additionally, *audience* takes value 1 when the corresponding experimental condition applies and 0 otherwise. Finally, *fraudster* identifies a dummy variable that is equal to 1 when the individual does not hold a validated ticket or pass and 0 otherwise. In model (2), we add the *inspection x fraudster* and *audience x fraudster* variables, representing interaction terms between *fraudster* and the *inspection* condition.

In model (3), the following control variables are also included in the analysis. First, based on the actors' evaluation given by the experimental subjects in the laboratory during the casting phase (see footnote 5), we categorise actors and actresses depending on their relative score (high or low) and include them as dummies in the regression, with the high score actress taken as the baseline category. Second, with respect to passengers' socio-demographics, we control for apparent age (coded as a continuous variable), and we include dummy variables for gender, ethnicity (identifying Caucasian—the baseline group—Arab, African, Asian or any other ethnic group), apparent wealth (poor—the baseline group—average and rich) and whether or not religious signifiers were visible. We also controlled for environmental conditions. The geolocation is captured through three dummy variables (Centre Metropolitan Lyon—the baseline category—North-East Metropolitan Lyon and South-East Metropolitan Lyon). Fixed effects are added for the transport line and for the time of day. More specifically, we included a set of dummy variables for each of the main tram lines in our sample (T1, T2 and T4), with the remaining, minor lines representing the omitted reference category, as well as dummy variables for each time interval (morning from 9:00 a.m. to 11:59 a.m.—the baseline category—early afternoon from 12:00 p.m. to 2:59 p.m. and late afternoon from 3:00 p.m. to 6:15 p.m.).¹⁵ We also included dummies for whether the public vehicle was crowded and for the noticeability of the scene on the street. Weather was coded as a set of binary variables (sunny—the baseline category—cloudy and rainy). Finally, model (4) is similar to model (3), except that the subject's gender dummy and the actor dummies are replaced with three indicator variables capturing the gender composition of the actor-passenger pair (with female pairs as the baseline category). This aimed to test whether passengers reacted differently with someone more similar to them.

Model (1) shows that the average probability of accepting the banknote increases by 16 percentage points after an inspection and by 17 percentage points when not holding a ticket. In model (2), the positive coefficient of the variable *inspection* indicates that, when the passenger holds a ticket, being inspected by controllers significantly increases the probability of accepting the banknote. Additionally, the interaction term between the variables *inspection* and *fraudster* in model (2) was not significant ($p = 0.602$), confirming that ticket inspections increased the unethical behaviour of both fraudsters and non-fraudsters on the street and, thus, loss recovery cannot be the only explanation of these cross-context spillover effects. Models (3) and (4) confirmed these findings with very minor changes in the coefficients of the variables of interest.

¹⁴ Since the estimated coefficients on interactions in ordered models are difficult to interpret with standard marginal effects (Ai and Norton, 2003), we run the whole analysis with linear probability models.

¹⁵ The information about the time of day, the geolocation and the transport line were grouped into categories to preserve the information without over-parameterising the model. This avoided singleton dummies and too-sparse data in certain categories, and allowed a reasonable amount of variation among our key variables within each category.

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) | |
|-------------------------------------------------------------------------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
| | Coeff. | SE | Coeff. | SE | Coeff. | SE | Coeff. | SE |
| Inspection (baseline = no inspection) | 0.165*** | 0.045 | 0.188*** | 0.063 | 0.150*** | 0.065 | 0.160** | 0.066 |
| Audience (baseline = no audience) | -0.031 | 0.044 | 0.012 | 0.059 | 0.051 | 0.061 | 0.034 | 0.061 |
| Fraudster (baseline = no fraudster) | 0.166*** | 0.037 | 0.207*** | 0.058 | 0.177*** | 0.057 | 0.171*** | 0.058 |
| Inspection x fraudster | | | -0.047 | 0.090 | -0.012 | 0.087 | -0.021 | 0.089 |
| Audience x fraudster | | | -0.086 | 0.087 | -0.095 | 0.086 | -0.075 | 0.087 |
| Constant | 0.342*** | 0.034 | 0.321*** | 0.040 | 0.431*** | 0.122 | 0.545*** | 0.120 |
| <i>Actors/actress (baseline = higher - score actress)</i> | | | | | | | | |
| Lower-score actress | | | | | 0.189*** | 0.049 | | |
| Higher-score actor | | | | | -0.016 | 0.071 | | |
| Lower-score actor | | | | | 0.073 | 0.051 | | |
| <i>Gender interaction (baseline = female actress, female passenger)</i> | | | | | | | | |
| Female actress, male passenger | | | | | | | 0.029 | 0.046 |
| Male actor, female passenger | | | | | | | -0.049 | 0.057 |
| Male actor, male passenger | | | | | | | -0.028 | 0.050 |
| Male passenger (baseline = female passenger) | | | | | 0.023 | 0.037 | | |
| Age ^e | | | | | 0.028** | 0.014 | 0.024* | 0.014 |
| <i>Time of day (baseline = 9:00 a.m.-1:59 a.m.)</i> | | | | | | | | |
| 12:00 p.m.-2:59 p.m. | | | | | -0.071 | 0.044 | -0.059 | 0.045 |
| 12:00 p.m.-9:15 p.m. | | | | | -0.074 | 0.048 | -0.056 | 0.048 |
| <i>Geolocation (baseline = Centre Metropolitan Lyon)</i> | | | | | | | | |
| North-East Metropolitan Lyon | | | | | -0.030 | 0.047 | -0.027 | 0.048 |
| South-East Metropolitan Lyon | | | | | -0.070 | 0.051 | -0.056 | 0.051 |

Table 1. *Continued*

| | Model (1) | | Model (2) | | Model (3) | | Model (4) | |
|----------------------------------------------------------|-----------|--------|-----------|--------|-----------|--------|-----------|----|
| | Coeff. | SE | Coeff. | SE | Coeff. | SE | Coeff. | SE |
| <i>Dependent variable: decision to take the banknote</i> | | | | | | | | |
| <i>Line public transport (baseline = other)</i> | | | | | | | | |
| T1 | | | 0.028 | 0.072 | 0.034 | 0.073 | | |
| T2 | | | -0.044 | 0.085 | -0.094 | 0.085 | | |
| T4 | | | 0.011 | 0.087 | 0.032 | 0.089 | | |
| <i>Ethnicity (baseline = Caucasian)^e</i> | | | | | | | | |
| Arab | | | 0.062 | 0.055 | 0.038 | 0.054 | | |
| African | | | 0.063 | 0.051 | 0.051 | 0.051 | | |
| Asian | | | 0.079 | 0.117 | 0.078 | 0.125 | | |
| Other | | | 0.090 | 0.126 | 0.102 | 0.127 | | |
| <i>Social appearance (baseline = poor)^e</i> | | | | | | | | |
| Average | | | -0.204*** | 0.050 | -0.198*** | 0.050 | | |
| Rich | | | -0.254*** | 0.076 | -0.255*** | 0.078 | | |
| Religious signs (baseline = no religious signs) | | | -0.227** | 0.110 | -0.215** | 0.107 | | |
| Crowded (baseline = not crowded) | | | -0.048 | 0.038 | -0.061 | 0.038 | | |
| <i>Weather (baseline = sunny)</i> | | | | | | | | |
| Cloudy | | | -0.017 | 0.052 | -0.041 | 0.052 | | |
| Rainy | | | -0.086 | 0.080 | -0.151* | 0.082 | | |
| Someone could notice the scene (baseline = no one) | | | -0.041 | 0.039 | -0.045 | 0.039 | | |
| Observations | 708 | 708 | 708 | 708 | 708 | 708 | | |
| R ² | 0.055 | 0.056 | 0.15 | 0.15 | 0.3 | 0.3 | | |
| Prob > F | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | | |

Notes: Table 1 reports the coefficients from linear probability estimates as well as robust SE. ^e Estimated by the research assistant. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (Wald tests).

This was after controlling notably for time of day, geolocation and public transport line, which had no significant effect on the likelihood of taking the banknote.¹⁶ A few socio-demographic matters were noted: older subjects were more likely to take the banknote (possibly driven by a selection effect as, on average, wealthier older people use public transport less) while people of poorer appearance were also significantly more likely to violate the norm. Finally, individuals with visible religious signs exhibited a lower propensity to behave unethically in the second stage of our quasi-experiment.

To further check the robustness of our results, we replicated the analysis provided in model (3) and included either interaction terms between the time of day and each transport line category, or between geolocation and time of day (see model (1) and model (2), respectively, in Table E8 in Online Appendix E). In Table E9 in Online Appendix E, we instead included hourly dummies for the time of day, a dummy for each minor line, and local district fixed effects for geolocation. The magnitude and significance of the coefficients of interest remained almost the same as in Table 1. Furthermore, in Online Appendix A.4, we report the analysis based on propensity score matching, where we matched controlled and not controlled passengers on some key observable environmental variables. This undertook to account for the fact that ticket controls might not be entirely random but might rather vary depending on the time of day, geolocation and public transport line. The matching procedure allowed us to restrict the analysis to observations for which the controlled and not controlled subjects were more similar. The results of this analysis (reported in Tables E10 and E11 in Online Appendix E) confirmed the results reported in Table 1, suggesting that differences in timing and location of the inspections were unlikely to be responsible for our results.¹⁷

Overall, this analysis confirms that the enforcement of the deterrence institution reduces the intrinsic honesty not only of fare dodgers, but also of law-abiding passengers. To dig deeper into the mechanisms that could explain this result, we investigated whether the strength of the deterrence institution matters beyond the occurrence of a ticket inspection. This strength can be proxied by the size of the inspection teams.¹⁸ In Table 2, we focus on the inspection condition only, and isolate the impact of the number of ticket inspectors during a control on the decision to take the banknote, estimating a set of linear probability models. In model (1), the independent variables include one that indicates whether the subject is a fraudster, and a set of dummies for the inspection team size grouped into three categories (2–5—the baseline category—6–10 and 11–20). We also control for whether the inspection occurred at the bus stop or on board the public vehicle. In model (2), we add interaction terms between the set of dummies identifying the number of inspectors and the Fraudster variable. Finally, in model (3) we add the same individual and environmental controls, fixed effects for time of day, geolocation, and transport line as in Table 1.

Model (1) shows that the probability of taking the banknote is sensitive to the size of the inspection team, since the coefficients associated with a team of six to ten inspectors and with a

¹⁶ As already mentioned, 31% of the observations were collected in the presence of an experimenter who was not blind to the research questions. Controlling directly for this presence in the regression analysis does not change the results (see Table E7 in Online Appendix E).

¹⁷ Time of day, geolocation and public transport line may only be imperfect proxies of how ticket inspections are conducted in the field by the transport company. Selection on unobservables may still be present. Using the method developed by Oster (2019), we assessed whether unobserved characteristics that drive ticket controls could bias our estimates. If anything, we found that our results underestimate the true effect of ticket controls on the probability of accepting the 5-euro banknote.

¹⁸ Figure D3 in Online Appendix D shows a histogram with the distribution of the number of ticket inspectors per inspection in our quasi-experiment.

Table 2. *Effect of the Number of Ticket Inspectors on the Decision to Take the Banknote.*

| <i>Dependent variable: decision to take the banknote</i> | Model (1) | | Model (2) | | Model (3) | |
|-----------------------------------------------------------------------------------------------------------|------------------|-------|------------------|-------|------------------|-------|
| | Coeff. | SE | Coeff. | SE | Coeff. | SE |
| Fraudster (baseline = no fraudster) | 0.188*** | 0.071 | 0.413*** | 0.123 | 0.509*** | 0.115 |
| <i>Number of controllers (baseline = 1–5)</i> | | | | | | |
| 6–10 | 0.466** | 0.211 | 0.669*** | 0.268 | 0.688*** | 0.214 |
| 11–20 | 0.581*** | 0.216 | 0.918*** | 0.267 | 1.090*** | 0.247 |
| <i>Fraudster x number of controllers</i> | | | | | | |
| Fraudster x 6–10 controllers | | | –0.210 | 0.162 | –0.226 | 0.155 |
| Fraudster x 11–20 controllers | | | –0.543*** | 0.190 | –0.612*** | 0.220 |
| Control on board (baseline = control at the stop) | 0.384* | 0.214 | 0.450* | 0.233 | 0.658*** | 0.204 |
| Constant | 0.032 | 0.210 | –0.178 | 0.253 | –0.691** | 0.277 |
| <i>Actors/actress (baseline = higher – score actress)</i> | | | | | | |
| Lower-score actress | | | | | 0.305*** | 0.093 |
| Higher-score actor | | | | | –0.001 | 0.192 |
| Lower-score actor | | | | | 0.047 | 0.097 |
| <i>Additional controls for passengers (gender, age, ethnicity, social appearance, religious signs)</i> | No | | No | | Yes | |
| <i>Additional controls for the environment (weather, audience in the vehicle, audience on the street)</i> | No | | No | | Yes | |
| <i>Time of day dummies</i> | No | | No | | Yes | |
| <i>Geolocation dummies</i> | No | | No | | Yes | |
| <i>Line public transport dummies</i> | No | | No | | Yes | |
| Observations | 199 [§] | | 199 [§] | | 199 [§] | |
| R ² | 0.060 | | 0.097 | | 0.339 | |
| Prob > F | 0.014 | | 0.002 | | 0.0000 | |

Notes: Table 2 reports the coefficients and robust SEs from linear probability estimates on the inspection condition. [§] Five observations were excluded because the information about the number of inspectors was missing. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ (Wald tests).

team larger than ten inspectors are both positive and significant. In models (2) and (3), the two coefficients are also significant, suggesting that non-fraudsters, the reference category, were more likely to accept the banknote when inspected by a medium-sized or a large team of controllers rather than a small one. Moreover, the coefficient associated with a team of more than ten inspectors is larger than the coefficient associated with a team of six to ten inspectors, suggesting that non-fraudsters were more likely to accept the banknote as the team of inspectors increased in size ($p = 0.026$ in model (2), $p = 0.003$ in model (3)). It is interesting to note that in both models (2) and (3), the coefficient of the interaction term between being a fraudster and being inspected by a medium-sized team is not significant, indicating no difference between fraudsters and non-fraudsters with respect to the impact of a team with two to five inspectors versus a team with six to ten inspectors. In contrast, the interaction term between being a fraudster and being inspected by a large size team is significant and negative. This indicates that the non-fraudsters had the greater reaction to an increase in the size of the inspection teams. This finding can help to discern the various possible mechanisms triggering the spillover effects of the deterrence institution on intrinsic honesty, which we discuss in the next section. Finally, being controlled

on board, rather than at the bus/tram stop, significantly increased the probability of passengers' accepting the banknote.

This analysis supports our third result: the size of the inspection team increases the spillover effects of the deterrence institution on intrinsic honesty, especially for non-fraudsters.

3. Discussion

Our results provide strong evidence for cross-context spillover effects of inspections and sanctions on intrinsic honesty. Strikingly, these effects equally applied to fraudsters and to non-fraudsters. In what follows, we discuss which mechanisms could explain these spillover effects. We begin with the mechanisms that receive less support from our data and move towards those more consistent with our findings.

3.1. *Indirect Reciprocity*

Negative direct reciprocity against the authority that signals distrust by enforcing inspections is ruled out by design, since behaviour on the street cannot affect the transport company. However, 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 exploit a 'kind' third party who has just offered them money. Second, indirect reciprocity often arises for strategic motives (e.g., Engelmann and Fischbacher, 2009; Stanca, 2009) that are absent in our setting.

3.2. *Loss Recovery*

Fraudsters' willingness to recover a loss after being fined might conceivably explain their subsequent unethical conduct. However, it cannot explain the negative effect of inspections on non-fraudsters' intrinsic honesty across contexts. Moreover, we found no difference in the banknote acceptance rate between fraudsters who paid their fine on the spot ($N = 30/41 = 73.17\%$) and those who did not ($N = 33/52 = 63.46\%$; χ^2 test, $\chi^2(1) = 0.989$, $p = 0.320$). Since almost two-thirds of the fines that are not paid immediately are never recovered by the company, we know that a significant proportion of those who did not pay their fine on the spot will not actually suffer a monetary loss. Thus, loss recovery is unlikely the main mechanism behind our findings.

3.3. *Signalling*

An alternative mechanism could be that ticket inspections prompt people to update their belief about the intrinsic cost of honesty. For example, in their self-signalling model, Benabou and Tirole (2003) assumed that people have an imperfect self-knowledge. In our context, this may be an imperfect knowledge of their morality (how intrinsically costly it is to not validate the ticket). A passenger might interpret a ticket inspection as a signal that their intrinsic honesty is low ('*I am inspected because I am suspected of being dishonest*'), and revise their beliefs about the intrinsic cost to them. This might, in turn, affect a subsequent moral decision, inducing the person to accept the banknote more often in the following context. Benabou and Tirole's (2003) theory hinges on the assumption that the principal (the public transport company in our setting) possesses certain relevant information regarding the unknown characteristic of the agent or task

at hand. However, this is unlikely in our setting (inspectors target everyone in a vehicle, including law-abiding passengers).

A related signalling interpretation is that the inspection makes salient to a proportion of non-fraudsters that they paid for their ticket not because they are using a public service, but because they otherwise fear being fined. The inspection may signal their mainly extrinsic motivation, whose higher saliency might make them more willing to take the banknote in the new context where money can be earned unethically, but without the risk of sanction. However, it is unclear how and why this signalling mechanism would depend on the size of the inspection team, given our finding that the spillover effects on non-fraudsters were particularly reactive to the number of inspectors.

3.4. *Moral Balancing*

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 may lead to moral cleansing if the sanction reduces the dissonance between the individual's self-image and his desired moral self. However, as already noted, in acceptance of the banknote, there was no difference between the fraudsters who paid their fine on the spot—leading to possible moral cleansing—and those who did not. And again, there is no obvious link between such moral balancing strategies and the sensitivity to the size of the inspection team.

3.5. *Emotions*

Inspections might trigger negative emotions in both fraudsters and non-fraudsters; the former may feel anger or shame at being fined and the latter, sadness due to the experience of distrust, or anger when they observe rule violations. Cross-context spillovers might then emerge as a consequence of the passenger's mood that leads to their punishing whoever can be associated with the transport company (e.g., another passenger) or society in general (see, in other contexts, e.g., Card and Dahl, 2011; Munyo and Rossi, 2013). To investigate whether emotions arising during an inspection might affect passengers' subsequent misbehaviour, we explored the impact of a visible emotional reaction (crying, screaming) expressed during an inspection. This analysis revealed that the banknote acceptance rate among detected fraudsters did not depend on their expressing a strong emotion ($N = 11/16$, 68.75%) or not ($N = 50/74$, 67.57%) after being fined ($\chi^2(1) = 0.01$, $p = 0.927$). However, people may feel angry or sad without any overt expression.

We therefore explored whether ticket inspections made non-fraudsters susceptible to emotional responses by means of a new study. Indeed, this would be a precondition for emotions to be the general mechanism behind spillovers. Several weeks after the main experiment, we conducted a survey of 160 passengers who validated a ticket or a pass on public transport in Lyon, either after a ticket inspection (51 subjects) or without a ticket inspection (109 subjects). Following the same identification procedure as in our quasi-experiment, we asked passengers to self-report their feelings of happiness and nervousness using self-assessment manikins (Lang, 1980) (see Figure D8 in Online Appendix D). Self-reported happiness and nervousness after a ticket inspection ($N = 51$, mean = 3.88, SD = 1.16 and mean = 2.24, SD = 1.36, respectively, on a scale from 1 to 5) and when no inspection occurred ($N = 109$, mean = 4.14, SD = 0.92 and mean = 1.93, SD =

1.08, respectively) revealed that the inspected non-fraudsters were less happy and more nervous compared to uninspected ones. However, while reported emotions for the non-inspected tended to be closer to the mean than were those for the inspected passengers, the observed difference between these two groups was not statistically significant (Mann-Whitney test, $z = 1.13$, $p = 0.257$ for happiness; $z = -1.149$, $p = 0.251$ for nervousness).

Our data therefore provide little evidence that emotions triggered the spillover effects observed in our main experiment.¹⁹ We acknowledge, however, the limitation that the sample size of the inspected non-fraudsters in the survey is smaller, and the confidence band larger than that of the non-inspected non-fraudsters. Moreover, measurement errors could have played a role since we had only one question to measure each of the two emotions. Also, there was insufficient statistical power to detect any effect on emotions of the size of inspection teams. The occurrence of an inspection may also have generated a selection bias in the willingness to participate in our survey. Therefore, we cannot reject the role of emotions, especially for fraudsters, and suggest that more systematic investigations of this mechanism should be conducted.

3.6. Social Norms

Finally, the fact that individuals, and in particular non-fraudsters, reacted to the size of the inspection team suggests that the spillover effect might be driven not so much by the inspection itself, but rather by the information that is conveyed when many people are observed being inspected. The visibility of inspections may affect people's perception of the injunctive norm (what one ought to do or not do) or the descriptive norm (what most people do), decreasing one's intrinsic honesty as a consequence (Gino *et al.*, 2009). We investigated the first hypothesis by means of a second laboratory experiment ('laboratory experiment 2'). In this experiment, we elicited the injunctive social norm following the same procedure as in experiment 1, but *after* new subjects ($N = 96$) had played a simplified version of the public transport game of Dai *et al.* (2017) and received feedback on their payoff in this game (see details in Online Appendix B.2 and instructions in Online Appendix C.2). In this incentivised game, subjects had to decide whether or not to purchase a ticket, being uninformed of the exact probability of a control (50%). The results reject that perceptions of the injunctive norm differ between inspected and non-inspected non-fraudsters in this game.²⁰

Nevertheless, inspections might still inform people about the descriptive norm in the field (Sliwka, 2007; Dickinson *et al.*, 2015). If ticket inspections signal the prevalence of rule violations, people may revise downwards their perception of society's descriptive norm after an inspection. This may particularly affect those who had initially higher beliefs about the honesty of citizens (thus, presumably more the non-fraudsters than the fraudsters) and, as a result, it may weaken their own moral stance. This effect might be stronger if a larger inspection team is perceived as

¹⁹ Note that, even when emotions are measured by physiological responses, there is no consensus in the (limited) experimental literature on the relationship between emotions and unethical behaviour. In a cheating game with no risk of detection, Pittarello *et al.* (2018) found a correlation between a higher emotional arousal and a lower likelihood of cheating. In contrast, in a tax evasion game where fraud could be detected and fined, Coricelli *et al.* (2010) showed that cheaters tend to be more emotionally aroused than non-cheaters, both at the time of deciding whether to evade and in reaction to an audit; compliers were not more emotionally aroused after an audit than when not audited.

²⁰ 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 E12 in Online Appendix E.

a signal of a higher prevalence of violations. Our results in the field are consistent with such a normative channel: non-fraudsters were significantly more likely to accept the banknote when they had been inspected by a medium-sized team of inspectors compared to a small team, and even more so when inspected by a team composed of more than ten individuals than when inspected by a medium-sized team, whereas fraudsters reacted less to the size of the inspection team.²¹ While we cannot unambiguously isolate a unique explanation for the observed spillover effects, our evidence points in the direction that (the strength of) inspections might act as a proxy for the descriptive norm, spreading unethical behaviour across contexts.

4. Conclusion

Modern societies have developed centralised 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 such institutions in their context of application cannot isolate their pure effect on intrinsic honesty, since this is confounded by other factors such as material cost-benefit considerations (e.g., avoiding a sanction) or direct reciprocity. By studying their effect outside their scope of application, our quasi-experiment reveals that the relationship between deterrence institutions and intrinsic honesty is more complex than might be expected.

Deterrence institutions create incentives to behave honestly to avoid a sanction but, at the same time, as our evidence has shown, may also effect a reduction in intrinsic honesty. Instead of observing an educative effect across contexts, we found that, following a ticket inspection not only evaders but also those who abided by the law behaved unethically in a setting where the institution does not apply. The enforcement device, when made visible to individuals, might act as a proxy of the (otherwise less salient) prevailing descriptive norm, spreading unethical behaviour in contexts other than that directly targeted by the institution. Our results do not mean that such institutional enforcement is detrimental to compliance—our study is silent about the impact of ticket inspection on the willingness of passengers to pay for their next fares. Building on the contribution of Becker (1968), a huge theoretical and empirical literature has shown the positive effects of deterrence on compliance (see the review of Chalfin and McCrary, 2017). But our results point to the existence of a negative externality of this deterrence institution on intrinsic honesty, something that has largely been ignored both in the literature and by policymakers. Our results are consistent with those of Frey (1997) who showed that distrusting public laws can affect civic virtues negatively.

Teasing the mechanisms behind the negative spillovers from inspections on the level of intrinsic honesty of fraudsters and non-fraudsters requires additional investigation. Indeed, these mechanisms are not necessarily the same for fraudsters and non-fraudsters and they may not be

²¹ In the survey that we conducted several weeks after our experiment, we also elicited the beliefs of the respondents about the percentage of passengers travelling without a valid transport ticket/pass on the transport network in Lyon. We did not observe that a ticket inspection changes the beliefs of non-fraudsters in the field about the prevalence of fare evasion on public transport in Lyon. Non-inspected non-fraudsters: $N = 108$ (one missing observation), mean belief about the percentage of fraudsters = 30.93%; inspected non-fraudsters: $N = 50$ (one missing observation), mean belief = 29.28%. Mann-Whitney test, $z = 0.441$, $p = 0.659$. Insufficient variation in the data did not allow us to test an effect of the inspection team size. It is still possible that information about others' norm violations, conveyed by the number of inspectors, receives more attention only when it is relevant to the individual's goals than if it is not, so that the cross-context spillover spreads once people actually have the opportunity to behave unethically.

unique. In particular, it might be useful to elicit in a large-scale study how passengers' empirical norms about compliance varies with ticket inspections, with the size of inspection teams and with the number of non-compliers fined. This would help identify how the deterrence institution affects the perceived social norm and whether spillovers vary with such normative views. Since we cannot refute the role of emotions, perhaps especially of fraudsters who are publicly exposed as cheaters in the case of an institutional control, it would be important to induce emotions exogenously to measure the extent to which their variations affected spillovers. Another extension could be to introduce rewards (such as symbolic thank-you cards or bonuses on loyalty cards) given by inspectors to compliers to determine whether this would reduce the spillover for this group, which could be anticipated if self-signalling plays a role in the spillover.

Teasing out these mechanisms would help to refine the policy implications raised by our study. The negative spillover of making a deterrence policy implementation visible to individuals suggests that crackdown interventions should be used with parsimony if there is a willingness to limit negative externalities. If large inspection teams signal a high fraud rate and contribute to weakening rather than strengthening the compliance norm, inspections conducted by small teams of inspectors in plain clothes might generate less spillover effects across contexts, at least for non-fraudsters. If a self-signalling mechanism plays a role, the negative reactions of incentive-sensitive compliers to inspections might be counteracted by the introduction of positive incentives associated with inspections, such as loyalty card-type bonuses or any such expressions of approval. More generally, our study invites policymakers to adopt a broader view in evaluating the efficacy of an institution. A social welfare perspective requires ensuring that, in the aggregate, the positive effects of a deterrence institution are not cancelled out by spillovers into contexts beyond its direct target.

University of Lyon, France

Ca' Foscari University of Venice, Italy

University of Lyon, France

Additional Supporting Information may be found in the online version of this article:

Online Appendix Replication Package

References

- Ai, C. and Norton, E.C. (2003). 'Interaction terms in logit and probit models', *Economics Letters*, vol. 80(1), pp. 123–9.
- Andenaes, J. (1974). *Punishment and Deterrence*, Michigan University Press.
- Antonakis, J., d'Adda, G., Weber, R. and Zehnder, C. (2015). "'Just words? Just speeches?'" On the economic value of charismatic leadership', Working Paper, Department of Organizational Behavior, University of Lausanne.
- Ariely, D. (2012). *The Honest Truth About Dishonesty. How We Lie to Everyone—Especially Ourselves*, New York: Harper Collins.
- Axelrod, R. (1986). 'An evolutionary approach to norms', *American Political Science Review*, vol. 80(4), pp. 1095–111.
- Baldassarri, D. and Grossman, G. (2011). 'Centralized sanctioning and legitimate authority promote cooperation in humans', *Proceedings of the National Academy of Sciences*, vol. 108(27), pp. 11023–7.
- Becker, G.S. (1968). 'Crime and punishment: an economic approach', *Journal of Political Economy*, vol. 76(2), pp. 169–217.
- Belot, M. and Schröder, M. (2016). 'The spillover effects of monitoring: a field experiment', *Management Science*, vol. 62(1), pp. 37–45.
- Benabou, R. and Tirole, J. (2003). 'Intrinsic and extrinsic motivation', *Review of Economic Studies*, vol. 70(3), pp. 489–520.

- Benabou, R. and Tirole, J. (2006). 'Incentives and prosocial behavior', *American Economic Review*, vol. 96(5), pp. 1652–78.
- Bock, O., Baetge, I. and Nicklisch, A. (2014). 'Hroot: Hamburg Registration and Organization Online Tool', *European Economic Review*, vol. 71, pp. 117–20.
- Card, D. and Dahl, G.B. (2011). 'Family violence and football: the effect of unexpected emotional cues on violent behavior', *The Quarterly Journal of Economics*, vol. 126(1), pp. 103–43.
- Cassar, A., d'Adda, G. and Grosjean, P. (2014). 'Institutional quality, culture, and norms of cooperation: evidence from behavioral field experiments', *Journal of Law and Economics*, vol. 57(3), pp. 821–63.
- Chalfin, A. and McCrary, J. (2017). 'Criminal deterrence: a review of the literature', *Journal of Economic Literature*, vol. 55(1), pp. 5–48.
- Cohn, A., Maréchal, M.-A., Tannenbaum, D. and Zünd, C.L. (2019). 'Civic honesty around the globe', *Science*, 365(6448), pp. 70–3.
- Coricelli, G., Joffily, M., Montmarquette, C. and Villeval, M.-C. (2010). 'Cheating, emotions, and rationality: an experiment on tax evasion', *Experimental Economics*, vol. 13(2), pp. 226–47.
- Cour des Comptes. (2016). 'La lutte contre la fraude dans les transports urbains en Île-de-France: un échec collectif', *Rapport Public Annuel*, pp. 537–77.
- Dai, Z., Galeotti, F. and Villeval, M.-C. (2017). 'The efficiency of crackdowns: a lab-in-the-field experiment in public transportations', *Theory and Decision*, vol. 82(2), pp. 249–71.
- Dai, Z., Galeotti, F. and Villeval, M.-C. (2018). 'Cheating in the lab predicts fraud in the field. An experiment in public transportations', *Management Science*, vol. 64(3), pp. 1081–100.
- Dickinson, D.L., Dutcher, E.G. and Rodet, C.S. (2015). 'Observed punishment spillover effects: a laboratory investigation of behavior in a social dilemma', *Experimental Economics*, vol. 18(1), pp. 136–53.
- Dickinson, D.L. and Villeval, M.-C. (2008). 'Does monitoring decrease work effort? The complementarity between agency and crowding-out theories', *Games and Economic Behavior*, vol. 63(1), pp. 56–76.
- Di Tella, R. and Schargrodsky, E. (2004). 'Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack', *American Economic Review*, vol. 94(1), pp. 115–33.
- Effron, D.A. and Conway, P. (2015). 'When virtue leads to villainy: advances in research on moral self-licensing', *Current Opinion in Psychology*, vol. 6, pp. 32–5.
- Egu, O. and Bonnel, P. (2020). 'Can we estimate accurately fare evasion without a survey? Results from a data comparison approach in Lyon using fare collection data, fare inspection data and counting data', *Public Transport*, vol. 12(1), pp. 1–26.
- Engelmann, D. and Fichbacher, U. (2009). 'Indirect reciprocity and strategic reputation building in an experimental helping game', *Games and Economic Behavior*, vol. 67(2), pp. 399–407.
- Engl, F., Riedl, A. and Weber, R.A. (2017). 'Spillover effects of institutions on cooperative behavior, preferences, and beliefs', Working Paper 6504, CESifo, http://www.cesifo-group.de/DocDL/cesifo1_wp6504.pdf (last accessed: 3 August 2019).
- Falk, A. and Kosfeld, M. (2006). 'The hidden costs of control', *American Economic Review*, vol. 96(5), pp. 1611–30.
- Fehr, E. and Gächter, S. (2000). 'Cooperation and punishment in public goods experiments', *American Economic Review*, vol. 90(4), pp. 980–94.
- Fischer, P., Greitemeyer, T., Pollozek, F. and Frey, D. (2006). 'The unresponsive bystander: are bystanders more responsive in dangerous emergencies?', *European Journal of Social Psychology*, vol. 36(2), pp. 267–78.
- Fisman, R. and Miguel, E. (2007). 'Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets', *Journal of Political Economy*, vol. 115(6), pp. 1020–48.
- Frey, B. (1997). A Constitution for Knave Crowds Out Civic Virtues, *The Economic Journal*, vol. 107(443), pp. 1043–53.
- Frey, B.S. and Jegen, R. (2001). 'Motivation crowding theory', *Journal of Economic Surveys*, vol. 15(5), pp. 589–611.
- Gächter, S. and Schulz, J.F. (2016). 'Intrinsic honesty and the prevalence of rule violations across societies', *Nature*, vol. 531(7595), pp. 496–9.
- Galbiati, R., Henry, E. and Jacquemet, N. (2018). 'Dynamic effects of enforcement on cooperation', *Proceedings of the National Academy of Sciences*, vol. 115(49), pp. 12425–8.
- Garrett, N., Lazzaro, S., Ariely, D. and Sharot, T. (2016). 'The brain adapts to dishonesty', *Nature Neuroscience*, vol. 19(12), pp. 1727–32.
- Gino, F., Ayal, S. and Ariely, D. (2009). 'Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel', *Psychological Science*, vol. 20(3), pp. 393–8.
- Gintis, H. (2003). 'The hitchhiker's guide to altruism: gene-culture coevolution and the internalization of norms', *Journal of Theoretical Biology*, vol. 220(4), pp. 407–18.
- Gneezy, U., Kajackaite, A. and Sobel, J. (2018). 'Lying aversion and the size of the lie', *American Economic Review*, vol. 108(2), pp. 419–53.
- Gneezy, U. and Rustichini, A. (2000). 'Pay enough or don't pay at all', *Quarterly Journal of Economics*, vol. 115(3), pp. 791–810.
- Hampton, J. (1984). 'The moral education theory of punishment', *Philosophy and Public Affairs*, vol. 13(3), pp. 208–38.
- Hawkins, G. (1969). 'Punishment and deterrence: the educative, moralizing, and habituating effects', *Wisconsin Law Review*, vol. 1969, pp. 550–65.
- Keizer, K., Lindenberg, S. and Steg, L. (2008). 'The Spreading of Disorder', *Science*, vol. 322(5908), pp. 1681–5.

- Keolis. (2014). 'Fraude: comment lutter?', <https://fr.slideshare.net/Keolis/keo-fvri-2014> (last accessed: 23 September 2020).
- Knack, S. and Keefer, P. (1997). 'Does social capital have an economic payoff? A cross-country investigation', *Quarterly Journal of Economics*, vol. 112(4), pp. 1251–88.
- Lang, P. (1980). 'Behavioral treatment and bio-behavioral assessment: computer applications', in (J.B. Sidowski, J.H. Johnson and T.A. Williams, eds.), *Technology in Mental Health Care Delivery Systems*, pp. 119–37, Norwood: Ablex.
- Levitt, S.D. and List, J.A. (2007). 'What do laboratory experiments measuring social preferences tell us about the real world?', *Journal of Economic Perspectives*, vol. 21(2), pp. 153–74.
- List, J.A. (2011). 'Why economists should conduct field experiments and 14 tips for pulling one off', *Journal of Economic Perspectives*, vol. 25(3), pp. 3–16.
- Masclot, D., Noussair, C., Tucker, S. and Villeval, M.C. (2003). 'Monetary and non-monetary punishment in the voluntary contributions mechanism', *American Economic Review*, vol. 93(1), pp. 366–80.
- Mauro, P. (1995). 'Corruption and growth', *Quarterly Journal of Economics*, vol. 110(3), pp. 681–712.
- Munyo, I. and Rossi, M.A. (2013). 'Frustration, euphoria, and violent crime', *Journal of Economic Behavior & Organization*, vol. 89(C), pp. 136–42.
- Nisan, M. (1991). 'The moral balance model: theory and research extending our understanding of moral choice and deviation', in (W.M. Kurtine and J.L. Gewirtz, eds.), *Handbook of Moral Behavior and Development*, pp. 213–49, Hillsdale: Lawrence Erlbaum Associates, Inc..
- Nowak, M.A. and Sigmund, K. (2005). 'Evolution of indirect reciprocity', *Nature*, vol. 437(7063), pp. 1291–8.
- Oster, E. (2019). 'Unobservable selection and coefficient stability: theory and evidence', *Journal of Business & Economic Statistics*, vol. 37(2), pp. 187–204.
- Pentico, D.W. (1981). 'On the determination and use of optimal sample sizes for estimating the difference in means', *The American Statistician*, 35(1), 40–42.
- Peysakhovich, A. and Rand, D.G. (2016). 'Habits of virtue: creating norms of cooperation and defection in the laboratory', *Management Science*, vol. 62(3), pp. 631–47.
- Pittarello, A., Conte, B., Caserotti, M., Scriminand, S. and Rubaltelli, E. (2018). 'Emotional intelligence buffers the effect of physiological arousal on dishonesty', *Psychonomic Bulletin Review*, vol. 25(1), pp. 440–6.
- Sands, M.L. (2017). 'Exposure to inequality affects support for redistribution', *Proceedings of the National Academy of Sciences*, vol. 114(4), pp. 663–8.
- Shalvi, S., Dana, J., Handgraaf, M.J.J. and De Dreu, C.K.W. (2011). 'Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior', *Organizational Behavior and Human Decision Processes*, vol. 115(2011), pp. 181–90.
- Sharma, E., Mazar, N., Alter, A.L. and Ariely, D. (2014). 'Financial deprivation selectively shifts moral standards and compromises moral decisions', *Organizational Behavior and Human Decision Processes*, vol. 123(2), pp. 90–100.
- Sliwka, D. (2007). 'Trust as a signal of a social norm and the hidden costs of incentive schemes', *American Economic Review*, vol. 97(3), pp. 999–1012.
- Stanca, L. (2009). 'Measuring indirect reciprocity: whose back do we scratch?', *Journal of Economic Psychology*, vol. 30(2), pp. 190–202.
- Swami, V., Chan, F., Wong, V., Furnham, A. and Tovée, M.J. (2008). 'Weight-based discrimination in occupational hiring and helping behavior', *Journal of Applied Social Psychology*, vol. 38(4), pp. 968–81.
- Wang, L., Northcraft, G.B. and Van Kleef, G.A. (2012). 'Beyond negotiated outcomes: the hidden costs of anger expression in dyadic negotiation', *Organizational Behavior and Human Decision Processes*, vol. 119(1), pp. 54–63.
- Weisburd, D., Wyckoff, L.A., Ready, J., Eck, J.E., Hinkle, J.C. and Gajewski, F. (2006). 'Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits', *Criminology*, vol. 44(3), 549–92.
- Welsh, D., Ordóñez, L.D., Snyder, D.G. and Christian, M.S. (2015). 'The slippery slope: how small ethical transgressions pave the way for larger future transgressions', *Journal of Applied Psychology*, vol. 100(1), pp. 114–27.
- Wilson, J.Q. and Kelling, G.L. (1982). 'Broken windows: the police and neighborhood safety', *Atlantic Monthly*, vol. 249(3), pp. 29–38.
- Winter, F. and Zhan, N. (2018). 'Social norm enforcement in ethnically diverse communities', *Proceedings of the National Academy of Sciences*, vol. 115(11), pp. 2722–7.