

## TITLE PAGE

**Classification:** SOCIAL SCIENCES, Economic Sciences

**Title:** Fraud Deterrence Institutions Decrease Intrinsic Honesty

**Short Title:** Fraud Deterrence Institutions Decrease Honesty

**Authors:** Fabio Galeotti<sup>a</sup>, Valeria Maggian<sup>b</sup>, Marie Claire Villeval<sup>c,d</sup>

Formattato: Italiano (Italia)

### Affiliations:

<sup>a</sup> Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France.

<sup>b</sup> Ca` Foscari University, Department of Economics. Cannaregio 873, Fondamenta San Giobbe, 30121 Venice, Italy.

<sup>c</sup> Corresponding author: Univ Lyon, CNRS, GATE, UMR 5824, F-69130 Ecully, France.

<sup>d</sup> IZA, Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany

**Abstract:** Modern societies have developed formal institutions to discourage rule violations. The success of these institutions is measured with respect to their deterrence effect on the targeted misbehavior. Whether these institutions also affect socially desirable behavior beyond their immediate scope is usually ignored. Their overall effectiveness, however, crucially depends on this. Using a field experiment, we look at possible spillover effects, across contexts, from inspecting and sanctioning people on one of the most common unethical traits of human beings: dishonesty. We first identified fraudsters (n=350) and non-fraudsters (n=358) in public transport. Conditions vary whether these passengers' tickets were inspected or not by the transport company. Second, we measured intrinsic honesty: on the street, the same fraudsters and non-fraudsters were asked by actors if they lost a banknote that did not belong to them. The deterrence institution has no educative effect across contexts. To the contrary, ticket inspection fosters unethical behavior of *both* fraudsters and non-fraudsters in the new context and the effect is stronger when the number of inspectors is larger. Inspections crowd-out intrinsic honesty in contexts where people know that they cannot be enforced. If weak institutions impair intrinsic honesty in a society (Gächter and Schulz, 2016), our findings suggest that even strong deterrence institutions can weaken intrinsic honesty beyond their direct limits of application.

**Keywords:** Fraud deterrence Institutions, Intrinsic honesty, Spillovers, Field Experiment.

**Significance Statement:** 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 field experiment, we show that they affect intrinsic honesty outside their context of application. We identified fraudsters and non-fraudsters in public transport who were or not exposed to ticket inspections by the transport company. We then measured their intrinsic honesty: on the street, the same fraudsters and non-fraudsters were asked by actors if they lost a banknote that did not belong to them. Instead of having an educative effect, the deterrence institution increases unethical behavior of both fraudsters and non-fraudsters, crowding out intrinsic honesty in contexts not related to the institution.

## Introduction

Honesty and norm compliance are fundamental for the maintenance of trust and thus, for the development of prosperous societies (1, 2). Norms can be sustained by an internalization mechanism that makes individuals comply even in absence of any threat of punishment (3, 4). However, intrinsic honesty is not sufficient to prevent violations and it varies widely across cultures (5, 6). While peer punishment of violations facilitates norm compliance (7, 8), inspections and sanctions by centralized authorities are the most common deterrence institutions in modern societies. When they raise the costs of breaking the rule above its benefits (9), these institutions can discourage the targeted misbehavior (10-13). However, they sometimes crowd out the intrinsic motivation to comply (14-17), with spillovers onto adjacent activities (e.g., controlling employees' performance may reduce their punctuality) (18). This results from control averse individuals who reciprocate against a distrusting authority that reduces their freedom of choice. While the estimations of deterrence effects focus almost exclusively on the targeted behavior, indirect effects may expand across contexts, if people extrapolate what they learn in a domain to another domain (19). For example, past exposure to institutions fostering pro-social norms improves future pro-sociality and beliefs about others' cooperativeness even when the institution is no longer enforced (20, 21). Here, we investigate the existence and direction of cross-contexts spillover effects of deterrence institutions on intrinsic honesty.

On the one hand, experiencing a deterrence institution in the past may act as an educative tool for the future, fostering moral behavior beyond its direct scope of intervention. On the other hand, a deterrence institution may negatively affect intrinsic honesty across contexts if individuals evaluate their moral activities dynamically (22), if the institution signals the prevalence of weak moral norms in the society (23, 24) or if extrinsic incentives to behave morally crowd-out intrinsic motivation (25, 26).

To identify the existence and the direction of spillover effects of deterrence institutions on intrinsic honesty across contexts, we ran a natural field experiment. The experiment consisted of two stages and was conducted by teams composed of a research assistant and a professional actor (see Fig. S1). The first stage aims at identifying dishonesty in a natural setting where formal deterrence institutions can be 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, even if a ticket is valid for an hour. 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 few passengers boarding and validating or not their ticket. 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. In the Inspection condition (I), the targeted passenger was controlled by a team of ticket inspectors from the transport company during or at the end of the ride. In the No-Inspection condition (NI), no ticket inspection occurred.

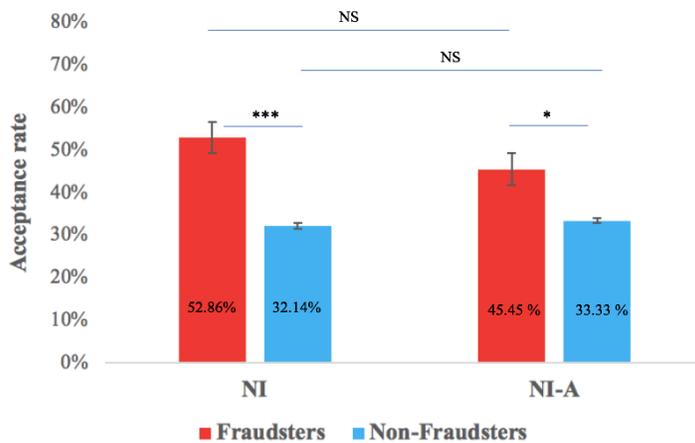
The second stage of the experiment was conducted on the street, where we measured each of the previously targeted passenger's intrinsic honesty in a context where no formal institution applies. The actor, while having a fake phone conversation to minimize interactions, suddenly bent down to seemingly 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 collected data regarding observable characteristics of the passenger and the environment. The actor's fake phone conversation was recorded to verify, ex-post, that the actor played the scene similarly across conditions (see Supplementary Methods). 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 condition (NI-A). 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.

This design allows us to correlate the honesty of the same passengers in the bus and on the street and to identify the causal effect of ticket inspection on the latter. In total, we collected 708 observations from 358 non-fraudsters and 350 fraudsters (see Supplementary Materials). The experiment constitutes a 3×2 factorial design with the first factor representing the three conditions described above (I, NI, NI-A) and the second factor the regular or irregular situation of the passenger on board the bus or tram.

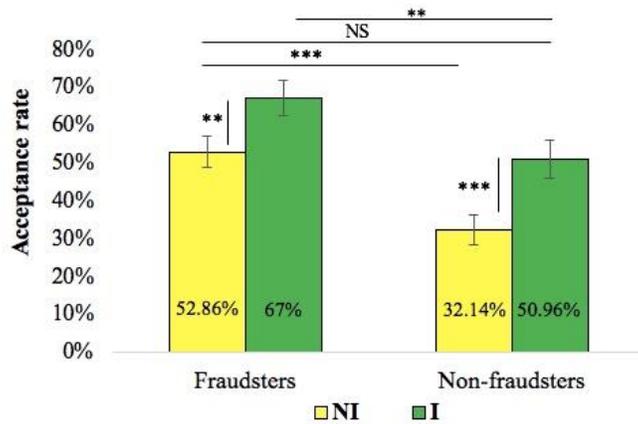
### Results

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 of the experiment, passengers without a validated ticket or pass (i.e., fraudsters) were more likely than passengers with a valid ticket (i.e., non-fraudsters) to claim ownership of the banknote on the street in the second stage of the experiment. Figure 1 presents the percentage of fraudsters and non-fraudsters who accepted the banknote in the NI and NI-A conditions. In NI, 52.86% of the fraudsters accepted the banknote compared to 32.14% of the non-fraudsters. This difference is significant (two-sided Chi-squared test:  $\chi^2(1) = 12.29$ ,  $P < 0.001$ ), 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. Although fraudsters were slightly less likely to accept 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 accepted the banknote in NI-A (33.33%) is not different from NI ( $\chi^2(1) = 0.04$ ,  $P = 0.841$ ).



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

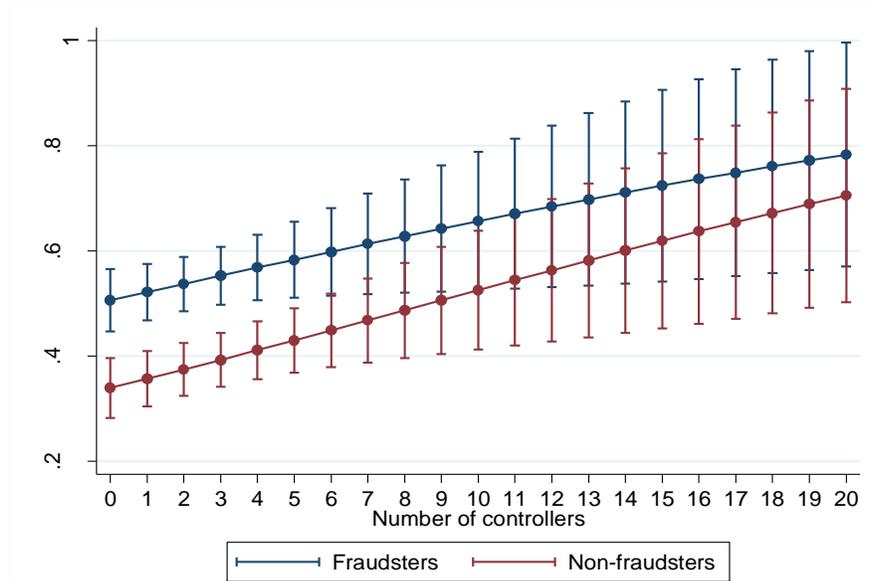
We found evidence that the enforcement of the deterrence institution in the first stage of the experiment spills over on the intrinsic honesty of fraudsters in the second stage (i.e., behavior on the street). Figure 2 displays the percentage of fraudsters and non-fraudsters who accepted the banknote,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



**Fig. 2.** Percentage of fraudsters and non-fraudsters accepting the banknote in the conditions with and without ticket inspection.

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 (4). To examine this, we consider the behavior of non-fraudsters after a ticket inspection. If loss recovery is 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 inspection ( $\chi^2(1) = 8.79$ ,  $P = 0.003$ ). The enforcement of the deterrence institution also reduced the intrinsic honesty of the law-abiding passengers. The percentage of passengers who accepted 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 ( $\chi^2(1) = 0.08$ ,  $P = 0.769$ ). A differences-in-differences regression analysis confirms that the spillover effect is of the same magnitude for fraudsters and non-fraudsters (see Supplementary Methods). Loss recovery cannot explain this cross-context spillover effect.

Finally, the probability to accept 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 a logit regression where the dependent variable is a dummy equal to 1 if a subject accepted the banknote (see Table S2). Thus, it is not only the enforcement of the deterrence institution that affects intrinsic honesty across contexts, but also its intensity.



**Fig. 3.** Probability to accept the banknote on the street depending on the intensity of ticket inspection in public transportation.

Logit regressions on the banknote acceptance decision, controlling for covariates characterizing the environment and the socio-demographics, confirm the robustness of our findings (see Table S2). The environmental characteristics (e.g., number of passengers, bus vs. tram) and most socio-demographics are insignificant, except for age that plays positively, and for apparent wealth and religious signs that play negatively.

**Discussion**

The negative effect of inspections on non-fraudsters’ intrinsic honesty across contexts is puzzling. We reject a number of explanations. Direct negative reciprocity against the authority that signals distrust by enforcing inspections is ruled out since behavior on the street cannot affect the transport company. We also reject explanations based on emotions. 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 in public transport several weeks after the experiment reveals that non-fraudsters’ self-reported happiness and emotional arousal 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,  $z = 1.13$ , two sided  $P = 0.257$  for happiness;  $z = -1.149$ , two sided  $P = 0.251$  for arousal) (see Supplementary Materials).

Complementary investigations do not support a normative explanation either, as inspections seem to affect neither injunctive norms (what one ought to do or not to do), nor descriptive norms (what most people do). To measure injunctive norms, we ran two additional experiments in the laboratory. Laboratory Experiment 1 ( $N = 30$ ) elicited whether accepting a banknote found by another person violates a social norm, by incentivizing subjects to guess the modal response of the other participants (see Supplementary Materials). Claiming ownership of a banknote found on the ground was considered as somewhat or very socially appropriate by 80% of the subjects, whereas accepting the banknote when someone else has found it was judged as somewhat or very socially inappropriate by 100% of them (rank-sum test:  $z = -4.84$ , two-sided  $P < 0.001$ ). In Laboratory Experiment 2 ( $N = 96$ ), new subjects played first a public transport game in which they had to decide whether purchasing or not a ticket, being uninformed of the exact probability of a control (50%); then, we elicited the norm about accepting the banknote. We reject that the injunctive norm differs between inspected and non-inspected non-fraudsters in the lab (rank-sum test:  $z = 0.232$ , two-sided  $P = 0.816$ ). However, inspections might inform on the descriptive norm in the field (23, 24). 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, afterwards weakening their own morals when undeserved money is proposed. We conducted a survey in public transport to test for this hypothesis: our data show that ticket inspection does not change the beliefs of non-fraudsters 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,  $z = 0.441$ , two sided  $P = 0.659$ ).

Finally, a psychological explanation in terms of moral licensing (22) could apply to non-fraudsters if, after an inspection that reinforced their positive self-image, they loosen their moral code while maintaining their self-concept of honesty (27, 28). 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, the acceptance rate of the banknote does not differ between the fraudsters who paid their fine on the spot (63.5%) and those who did not (73.2%) ( $\chi^2(1) = 0.989$ ,  $P = 0.320$ ).

The negative spillover effect of inspections and sanctions on intrinsic honesty is more likely to result from the signaling value of the deterrence institution about people's true nature (26): not only fraudsters are reminded that they are dishonest, but a fraction of the non-fraudsters are reminded that they behaved honestly just because there could be a ticket inspection. 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.

Modern societies have developed centralized institutions to protect citizens and assets against dishonest behavior. Since the honesty norms prevailing in the environment, i.e., the frequency of violations, can compromise intrinsic honesty in a society (5), 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, we show that the relationship between 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, they reduce intrinsic honesty. Instead of having an educative effect across contexts, 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.

**Material and methods:** The data and code for the statistical analyses will be available at figshare. The data set is composed by 708 observations from the field experiment, by the observations of the survey and by the observations of the additional experiments run in the laboratory.

**Acknowledgments:** We thank J. Andreoni, M. Bigoni, G. Charness, R. Kranton, J. Van de Ven, and audiences at seminars (at the Universities of Aarhus, Bilkent, Birmingham, Innsbruck, Maastricht, Montpellier, Paris 1) 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) for useful comments. We are grateful to R. Sauter and the TCL company in Lyon for facilitating our access to the network. **Funding:** 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) is gratefully acknowledged. This project has also received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 661645.

#### References:

1. Mauro P (1995) Corruption and Growth. *Q J Econ* 110(3):681-712.
2. Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. *Q J Econ* 112(4):1251-1288.
3. Axelrod R (1986) An evolutionary approach to norms. *Am Polit Sci Rev* 80:1095–1011.
4. Gintis H (2003) The hitchhiker’s guide to altruism: Gene-culture coevolution and the internalization of norms. *J Theor Biol* 220:407-418.
5. Gächter S, Schulz J-F (2016) Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531:496-499.
6. Cohn A, Fehr E, Maréchal M-A (2014) Business culture and dishonesty in the banking industry. *Nature* 516: 86–89.
7. Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Amer Econ Rev* 90(4): 980-994.
8. Masclet D, Noussair C, Tucker C, Villeval M-C (2003) Monetary and non-monetary punishment in the voluntary contributions mechanism. *Amer Econ Rev* 93:366-380.
9. Becker G-S (1968) Crime and punishment: an economic approach. *J. Polit. Econ.* 76(2):169-217.

10. Baldassarri D, Grossman G (2011) Centralized sanctioning and legitimate authority promote cooperation in humans. *P Natl Acad Sci USA* 108(27):11023-11027.
11. Di Tella R, Schargrodsky E (2004) Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *Amer Econ Rev* 94(1):115-133.
12. Ariely D (2012) *The honest truth about dishonesty. How we lie to everyone---especially ourselves* (Harper Collins, New York).
13. Gino F (2015) Understanding ordinary unethical behavior: why people who value morality act immorally. *Curr Opin Behav Sci* 3:107-111.
14. Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. *Q J Econ* 115(3):791-810.
15. Frey B-S, Jegen R (2001) Motivation crowding theory. *J Econ Surv* 15:589-611.
16. Falk A, Kosfeld M (2006) The hidden costs of control. *Amer Econ Rev* 96(5):1611-1630.
17. Dickinson D-L, Villeval M-C (2008) Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Game Econ Behav* 63:56-76.
18. Belot M, Schröder M (2016) The spillover effects of monitoring: a field experiment. *Manage Sci* 62(1):37-45.
19. Rand D-G, Peysakhovich A, Kraft-Todd G-T, Newman G-E, Wurzbacher O, Nowak M-A, Green J-D (2014) Social heuristics shape intuitive cooperation. *Nature Commun* 5, Article 3677.
20. Peysakhovich A, Rand D-G (2016) Habits of virtue: creating norms of cooperation and defection in the laboratory. *Manage Sci* 62(3):631-647.
21. Engl F, Riedl A, Weber R-A (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](http://www.cesifo-group.de/DocDL/cesifo1_wp6504.pdf) on 28 August 2018.
22. Nisan M (1991) The moral balance model: Theory and research extending our understanding of moral choice and deviation. *Handbook of moral behavior and development*, eds. Kurtine W-M, Gewirtz J-L (Lawrence Erlbaum Associates, Inc.), pp. 213-249.
23. Sliwka D (2007) Trust as a signal of a social norm and the hidden costs of incentive schemes. *Amer Econ Rev* 97(3):999-1012.
24. Dickinson D-L, Dutcher E-G, Rodet C-S (2015) Observed punishment spillover effects: a laboratory investigation of behavior in a social dilemma. *Exp Econ* 18:136-153.
25. Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: a theory of self-concept maintenance. *J Marketing Res* 45:633-644.
26. Benabou R, Tirole J (2003) Intrinsic and extrinsic motivation. *Rev Econ Stud* 70(3):489-520.
27. Benabou R, Tirole J (2006) Incentives and prosocial behavior. *Amer Econ Rev* 96(5):1652-1678.
28. Shalvi S, Dana J, Handgraaf M-J-J, De Dreu C-K-W (2011) Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ Behav Hum Decis Process* 115:181-190.

## Figure Legends

Figure 1. The figure depicts the percentage of fraudsters (red bars) and non-fraudsters (blue bars) in public transportation that accepted the banknote on the street in the two conditions without prior ticket inspection (NI and NI-A). The number of observations is respectively: N = 140 (NI, fraudsters), 140 (NI, non-fraudsters), 110 (NI-A, fraudsters), and 114 (NI-A, non-fraudsters). Audience has a significant effect neither on fraudsters nor on non-fraudsters. Error bars, mean  $\pm$  SEM. \* P < 0.1, \*\*\* P < 0.01, NS not significant, two-sided Chi-squared tests.

Figure 2. The figure depicts the percentage of fraudsters and non-fraudsters in public transportation that accepted the banknote on the street in the absence of an audience in the two conditions with and without prior ticket inspection (yellow bars for NI and green bars for I). The number of observations is respectively: N = 140 (NI, fraudsters), 100 (I, fraudsters), 140 (NI, non-fraudsters), and 104 (I, non-fraudsters). All pairwise comparisons are significant except non-fraudsters in I vs. fraudsters in NI. Error bars, mean  $\pm$  SEM. \*\* P < 0.05, \*\*\* P < 0.01, NS not significant, two-sided Chi-squared tests.

Figure 3. The figure depicts the probability of an individual to accept the banknote on the street depending on the number of inspectors who conducted the ticket control in public transportation (blue line for fraudsters and red line for non-fraudsters). The probability is computed from a logit regression where the dependent variable is a dummy equal to 1 if the passenger accepted the banknote and 0 otherwise (see Supplementary Methods). The number of observations is: N = 65 for [1, 5), 90 for [5, 10), and 45 for [10, 20]. Four observations were dropped because information about the number of inspectors was missing. The figure shows that the intensity of the inspection decreased intrinsic ethics in both fraudsters and non-fraudsters.