Public Choice (2023) 196:229–255
https://doi.org/10.1007/s11127-022-01000-5

Fines as enforcers' rewards or as a transfer to society at large? Evidence on deterrence and enforcement implications

Florian Baumann[1] · Sophie Bienenstock[2] · Tim Friehe[3] · Maiva Ropaul[4]

Received: 6 July 2022 / Accepted: 24 August 2022 / Published online: 21 October 2022
© The Author(s) 2022

Abstract
We analyze experimental data to assess whether the deterrent effect of expected fines depends on who receives the fines' proceeds. We compare behavior in treatments when the revenue is a reward for enforcement agents to the alternative when fines are transferred to society at large. Most important, with a fixed detection probability, potential offenders' material incentives are held constant across treatments. Our evidence suggests that the deterrent effect of expected fines is greater when enforcement agents obtain the fine revenue. Our results also document that the characteristics of enforcers who are willing to incur private costs to create a positive detection probability seem to depend on whether fines reward enforcers or are transferred to society at large.

Keywords Crime · Enforcement · Compensation · Experiment

JEL Classification C91 · D92 · K42

1 Introduction

The use of financial rewards to incentivize individual law enforcement agents was once a common practice in the United States but was largely eliminated by the late nineteenth century over concerns that such incentives distort enforcement effort (e.g., Parrillo, 2013).

✉ Tim Friehe
tim.friehe@uni-marburg.de

Florian Baumann
florian.baumann@zew.de

Sophie Bienenstock
sophie.bienenstock@univ-paris1.fr

Maiva Ropaul
maiva.ropaul@parisdescartes.fr

1 ZEW Mannheim, L 7,1, 68161 Mannheim, Germany

2 Université Paris I, Paris, France

3 Public Economics Group, University of Marburg, Am Plan 2, 35037 Marburg, Germany

4 Université de Paris, LIRAES, Paris, France

In some arenas, such as asset forfeiture laws, financial incentives remain commonplace in the United States and increasingly are the subject of scrutiny (e.g., Carpenter et al., 2015; Miceli & Johnson, 2016). Discussion regarding enforcement agents' financial incentives concentrates on both the level and allocation of enforcement resources. The relevant literature relates to the issue of deterrence only insofar as the induced change in enforcement effort lowers or increases expected sanctions. However, the possibility that financial rewards for law enforcement agents may affect deterrence even when the expected sanction remains the same has been neglected in the discussion thus far. The discussion's focus may be attributable to data availability. With field data, it is impossible to disentangle the deterrence implications of financial rewards resulting solely from the remuneration scheme (that is, with fixed enforcement effort) and the effects resulting from varied enforcement efforts (i.e., variation in the expected sanction).

This paper contributes to filling the gap by analyzing data from a laboratory experiment with varying enforcement agent remuneration schemes. By comparing offender choices at predetermined levels of enforcement effort, we can measure the deterrence effects resulting exclusively from the remuneration scheme while holding the expected sanction constant. The experimental methodology enables us to conduct a clean ceteris paribus comparison between compliance choices in a treatment in which fine revenue benefits society at large and enforcers receive a lump sum payment (where, in the lab, a charity serves as a proxy for society) and decisions in a treatment in which enforcers receive fines as rewards for effort.

In our experiment, we consider a taking game with three players. First, enforcers decide whether to incur a cost creating a positive detection probability or not. Next, potential offenders learn the level of the detection probability (either 0% or 50%) and then choose whether to take points from the potential victim. That sequence enables us to compare potential offenders' taking decisions across treatments, holding both the detection probability and the level of the sanction constant (thus only varying who receives the fine revenue).[1] By comparing offenders' decisions for a given level of enforcement effort (i.e., for a given expected sanction), we provide a clean test of whether potential offenders' compliance with the law depends directly on the enforcement agents' remuneration scheme. Three treatments are considered: in treatment FLAT, the potential fine revenue is allocated to society at large, and enforcers receive a lump sum payment; in treatment REWARD, the fine is transferred to the enforcer responsible for the offender's detection. We also consider a third treatment entitled CORRUPTION, in which the fine is intended to benefit society but can be diverted by the enforcer for a private gain (which may be interpreted as embezzlement). In treatment CORRUPTION, potential offenders condition their taking choice on the enforcer's decisions regarding both enforcement and possible fine diversion. That treatment allows us to shed light on the implications of receivers of the fine when the allocation is determined endogenously by the enforcer. As a theoretical basis for our experiment, we augment a standard model of law enforcement with risk-neutral potential offenders by accounting for offenders' inequity aversion or the possibility that the enforcers' decision conveys information about social norms to potential offenders.

[1] If enforcers and potential offenders had to decide simultaneously, potential offenders would base their violation decisions on their beliefs about the level of enforcement. Those beliefs would be treatment-specific because the enforcer only has material incentives to invest in enforcement in the treatment that rewards enforcers. Our sequential structure avoids that problem.

Our paper tests a central assumption of the standard model of optimal law enforcement, according to which a fine is considered to be a costless transfer from the offender to society, and the precise allocation of fine revenue is assumed to be irrelevant to potential offenders' compliance choices (see, e.g., Becker, 1968; Polinsky & Shavell, 2007). That *irrelevance* assumption also is adopted in very recent contributions that focus on the choice of enforcers' remuneration. For example, in the political economy model of Yahagi (2021), a democratic government determines the extent to which fine revenue is used in order to incentivize enforcement agents to invest in higher detection probabilities. In that paper, offender behavior is guided exclusively by the expected sanction. Likewise, Adamson and Rentschler (2021) discuss the allocation of enforcement resources to patrolling (higher detection probability) and investigating (lower probability of false convictions), assuming that potential offenders' decisions are guided by expected sanctions alone. Our analysis is inspired by Becker and Stigler (1974) who provide the first economic analysis of a remuneration regime in which enforcers are compensated from fines the detected offenders paid. They discussed, inter alia, the possible effects of the remuneration scheme on law enforcement effort but also that deterrence might be affected by non-effort factors like the potential effect of financial incentives on enforcers' honesty. At the time, Becker and Stigler's recommendation contrasted starkly with the common practice of paying public enforcement agents a flat salary. Those two scenarios, fines as enforcer rewards and fines as a transfer to society while enforcers receive a flat payment, represent our main experimental treatments.

Our analysis suggests that potential offenders are influenced by who receives fine revenue, a mechanism that is not taken into consideration in the standard model of optimal law enforcement. In our data, the enforcer's creation of a positive detection probability generates a weakly greater deterrent effect when enforcers benefit from fine revenue than when society benefits from fines. In terms of mechanisms, our finding is consistent with inequity aversion on behalf of potential offenders (e.g., Fehr & Schmidt, 1999). In contrast, the idea that enforcers who invest in detection probabilities without explicit financial motives signal to potential offenders strong social norms regarding the unacceptability of taking and, thereby, influence deterrence, is not supported by our evidence. Although our conclusions are based only on a laboratory experiment and we cannot identify a specific mechanism to account for our results, we believe that the findings are very relevant and promising for the study of optimal law enforcement.

In addition to our interest in enforcement's deterrent effect, we explore how the different regimes influence enforcement effort. Specifically, we study the frequency of costly law enforcement and the types of individuals engaging in it.[2] With respect to enforcement effort, our evidence shows that financial incentives induce more enforcers to invest into positive detection probabilities, corresponding to economic intuition. Moreover, the characteristics of enforcement agents who invest in the detection probability seem to depend on who receives fine revenue. More specifically, we find that the increase in the enforcement probability in response to financial incentives is most pronounced for individuals with low morality scores.[3] That result sheds light on how financial incentives influence the selection of law enforcement actors.

[2] Our results can be relevant to contributions such as Dharmapala et al. (2016) and Dickinson et al. (2015) that deal with norm enforcement incentives and self-selection of police officers.

[3] The morality score relied on in our empirical analyses results from subjects' self-reported attitudes regarding different norm violations (e.g., drunk driving) and originates from Traxler and Winter (2012).

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 introduces the experimental design, and Sect. 4 presents behavioral predictions. In Sect. 5, we present our empirical results. We offer some concluding remarks in Sect. 6.

2 Related literature

Our paper explores whether allocating fines to either reward enforcers or redistribute them to society at large influences law enforcement's deterrent effect and the frequency of enforcement. We thereby touch on contributions from both the experimental economics literature as well as the law and economics literature on private law enforcement.

Our experiment was stimulated by the mostly theoretical discussion of private law enforcement initiated by Becker and Stigler (1974) and Landes and Posner (1975). That literature asked, inter alia, whether enforcers should receive fine revenue, which is a prominent feature of private law enforcement (but could be applied in public law enforcement as well). Contributions to the strand of relevant research dealt extensively with topics such as the level of expected enforcement under private and public regimes (implying asymmetric detection probabilities). While Becker and Stigler (1974) argue that private enforcement can lead to the same enforcement level as public enforcement's theoretical optimum, Landes and Posner (1975) anticipate overenforcement under private legal arrangements. Later contributions include Polinsky (1980) and Garoupa (1997), who provide theoretical studies of the distortions likely to occur under private enforcement (e.g., over- or underenforcement) and their social implications. For example, Polinsky (1980) argues that public enforcement yields higher welfare than private enforcement in many circumstances. In all previous contributions, deterrence is determined solely by the level of the expected sanction. As such, our empirical work[4] explores whether the deterrent effect of a fixed expected sanction depends on the enforcer's remuneration, as would be consistent with inequity aversion or the possibility that the enforcer's decision conveys information about social norms to potential offenders. Our design allows us to isolate the effect of the enforcer's material incentives on potential offenders' behavior, *ceteris paribus*. To our knowledge, that possibility has not been considered in the relevant literature so far.

There is by now a rich empirical literature presenting evidence that incentives to raise revenue influence policing (e.g., Benson & Rasmussen, 1995; Garret & Wagner, 2009; Mast et al., 2000). For instance, Makowsky and Stratmann (2009, 2011) find that traffic citations increase when the municipality has pressing budgetary needs, thereby improving traffic safety. Makowsky et al. (2019) consider the joint effect of local budget deficits and possibilities for revenue retention from civil asset forfeitures on policing. The authors hint, for example, at the racial implications of a revenue motivation. From an analysis of forfeiture laws, Baicker and Jacobsen (2007) deduce that police behavior reacts to the incentives thereby created and that local governments reduce their transfers to the police by the amount collected through seizures. Along those same lines, the 2015 US Department

[4] Empirical evidence regarding the relationship of private and public law enforcement have been published. For example, Koyama (2014) provides a historical account of a system of private prosecution in England and why it came under strain during the Industrial Revolution. Focusing on bail jumping, Helland and Tabarrok (2004) compare the effectiveness of the different systems. Romaniuc et al. (2016), for instance, consider experimentally the interplay of public financial sanctions and private enforcement in the form of peer disapproval in a public-good game.

of Justice report "Investigation of the Ferguson Police Department" comes to the conclusion that "Ferguson's law enforcement practices are shaped by the City's focus on revenue rather than by public safety needs" (US Department of Justice, 2015, p. 2). We are interested herein whether the use of fine revenue *itself* has implications for deterrence and thus compare violation rates across treatments for the same level of enforcement effort.

The paper closest to ours in terms of research question and methodology is Xiao (2013). She considers third-party enforcers in a sender-receiver game who may have profit motives. In her game, third-party enforcers' payoffs increase unconditionally by punishing others. Therefore, enforcers with a profit motive may punish senders for dishonesty, independent of whether a violation actually occurred. In that specific context, Xiao (2013) finds a substantial degradation of punishment's ability to influence potential offenders' behavior. It is crucial that, in contrast to our setting, Xiao (2013) allows third-party enforcers to extort payments from compliant subjects. While we do not question the empirical relevance of that concern, we purposefully designed a setup without such framing incentives so as to provide clean evidence about the effect of enforcer incentives alone. Our evidence suggests that when the possibility of wrongful punishment is excluded, having fines that benefit enforcers can lead to a greater deterrent effect of enforcement.

Our paper also is related to other studies from the experimental literature on third-party punishment. For example, Leibbrandt and Lopez-Perez (2012) and Fehr and Fischbacher (2004) present evidence that third-party punishment often is motivated by inequity aversion. We focus on the behavior of potential offenders and find results that are consistent with inequity aversion as well.

Our research interest, the deterrence implications of enforcer remuneration schemes that go beyond their direct effects on enforcement effort, is related to the more general public choice literature on law enforcement. In that realm, Yahagi (2021) investigates how a democratic government chooses enforcers' remuneration. The work extends contributions that investigate law enforcement with self-interested principals, as in Garoupa and Klerman (2002) or Langlais and Obidzinski (2017). Regarding financial incentives, Friehe and Mungan (2021) employ a median-voter setup to show that the equilibrium enforcer liability for wrongful suspect stops likely will be suboptimal, that is, unable to curtail excessive policing. On a different note, Wickelgren (2003) explores the hypothesis that more costly forms of punishment are adopted to prevent policymakers from excessively punishing those agents with little political influence.

3 Experiment design, treatments, and procedures

Our computerized experiment consisted of six parts. Subjects first received a general introduction beforehand and then separate instructions for Parts 1–6. While subjects knew from the start that the experiment comprised several parts, the content of later parts was not revealed to participants until the appropriate time. We first summarize the experiment.

In Part 1, participants earn a fixed endowment in a real-effort task. Afterward, subjects learn their role in the experiment, either potential offender, potential victim, or enforcer.[5] In Parts 2–5, which constitute the core of our experiment, subjects participate in

[5] Neutral wording was adopted in the instructions. For example, roles were described as Player A, B, and C. See our Supplementary Material for a translation of the instructions given.

treatment-specific taking games with random matching in each part. In each taking game, enforcers first decide whether to create a positive detection probability at a private cost before potential offenders observe the enforcer's decision and decide about taking points from their potential victim. In Part 2, participants play one treatment in a direct-response format. In Part 3, participants play the *same* treatment, and potential offenders choose according to a strategy-method format.

In the strategy-method format, subjects make contingent decisions for all nodes at which they may have to play (e.g., Brandts & Charness, 2011). In our setting, potential offenders decide whether and how much to take from the potential victim conditional on all possible choices by the enforcer.[6]

Part 4 (5) mirrors Part 2 (3) for a different treatment. Potential victims remain passive throughout Parts 2–5. In Part 6, we collect information about our subjects' morality and their attitudes towards both justice and risk.

For each subject, one part out of Parts 2 to 5 was chosen randomly as payoff relevant. In addition, subjects received payoffs for the tasks in Part 6 (one of the questions was selected randomly to determine the payoff). Finally, a showup fee of 7 euros was paid to all participants. Throughout the experiment, payoffs are described as points; the exchange rate was 100 points = 1 euro. Subjects were informed of their earnings at the end of the experiment.

3.1 The key component: the taking game

Our taking game follows the experimental literature on the economics of crime (Rizzolli & Stanca, 2012; Schildberg-Hörisch & Strassmair, 2012). We depart from that literature by introducing a detection probability that depends on the enforcer's investment in effort.

At the beginning of the taking game, all participants in a group have 500 points in their accounts from Part 1's real-effort task.

First, the enforcer chooses whether to invest 50 points to create a detection probability amounting to 50%. Without investment, the detection probability is equal to zero.

Next, the potential offender learns the enforcer's choice and chooses how many points x to take from the potential victim's account, where $x \in [0, 500]$. The offender receives half of the points taken.[7] The symmetric endowments, the fact that the endowment was earned in a real-effort task, and the drastic inefficiency resulting from theft make not taking a social norm.

Finally, the offender's detection chance is determined. A detection can occur only if the enforcer created the 50% positive detection probability and the offender took points from the victim (our design does not allow for wrongful convictions, in contrast to Xiao, 2013). Detected offenders pay a fine amounting to 250 points but keep their criminal gains (as in, e.g., Rizzolli & Stanca, 2012). For simplicity, the fine is independent of the number of points taken; the fine amounting to 250 points applies when the offender has taken any positive number of points from the victim (as in, e.g., Schildberg-Hörisch & Strassmair, 2012).

[6] With the strategy method, offenders learn the enforcer's actual decision only after they have made their taking choices. However, as the taking choice that corresponds to the enforcer's decision is implemented, from a theoretical perspective, the game is similar to a game in which offenders observe the enforcer's choice before making their own decision.

[7] In reality, such inefficiency from theft can result from a victim's psychological cost (e.g., Falk and Fischbacher 2002).

Table 1 Treatments for the taking game

<table>
<tr><th>Treatment
FLAT</th><th>Treatment
REWARD</th><th>Treatment
CORRUPTION</th></tr>
<tr><td colspan="3">Part 1: Real-effort task</td></tr>
<tr><td colspan="3">Parts 2-5: The taking game</td></tr>
<tr><td colspan="3">Elicitation of incentivized beliefs</td></tr>
<tr><td colspan="2">Enforcer’s enforcement choice</td><td>Enforcer’s choice about enforcement <u>and</u> diverting the fine</td></tr>
<tr><td colspan="3">Enforcer’s choice announced to other players</td></tr>
<tr><td colspan="3">Potential offender chooses level of taking x from potential victim’s account, where $x \in [0,500]$, for a private gain of $\frac{x}{2}$</td></tr>
<tr><td colspan="3">Detected offenders pay fine amounting to 250 points</td></tr>
<tr><td>Fine goes to charity</td><td>Fine goes to enforcer</td><td>Fine goes to charity unless enforcer diverts it</td></tr>
<tr><td colspan="3">Part 6: Information about subjects’ characteristics</td></tr>
</table>

Before subjects choose their actions, potential offenders (enforcers) state their beliefs about enforcer (potential offender) conduct. Potential victims state their beliefs about both potential offenders’ and enforcers’ behavior. Players earn an additional 100 points if their stated beliefs roughly match either the share of people enforcing/taking or the average amount taken.[8]

[8] See our Supplementary Material for detailed instructions.

Before we proceed, we want to provide a motivation for our specific design of the taking game. Our research focuses on the implications of enforcer compensation on deterrence when the expected sanction is fixed across conditions. We thus seek to compare taking choices in circumstances in which the enforcer's compensation differs, but the enforcement effort and monetary fine are the same. To fulfill the requirement of a fixed expected sanction, we can compare only cases with exactly the same enforcer investment and sanction. The restriction to only two investment levels for enforcement effort allows us to collect enough observations for each circumstance. In addition, note that taking is a dominant strategy for potential offenders solely concerned about monetary payoffs. In other terms, we rely fully on the relevance of non-monetary aspects motivating behavior to create heterogeneity in terms of the taking choice.[9]

3.2 Treatments

We study three treatments denoted FLAT, REWARD, and CORRUPTION (see Table 1 for an overview).

In treatment FLAT, fines benefit society at large. It was common knowledge that we proxy society at large with the charity *Les Resto du Coeur*.[10] The enforcer's final payoff from the taking game is equal to the endowment minus any enforcement cost.

In treatment REWARD, fines benefit enforcers, meaning that the enforcer's expected remuneration depends on whether an offense is detected. The enforcer's final payoff from the taking game is equal to the endowment minus any enforcement cost plus the fine in case of detection.

Treatments FLAT and REWARD differ only regarding who benefits from fine revenues. The comparison between those two treatments is central in addressing our main research question concerning how the use of fine revenue influences law enforcement's deterrent effect.

In treatment CORRUPTION, fines are *meant* to benefit society at large; however, the enforcer can divert any collected fine. Accordingly, the enforcer makes two decisions, whether to enforce and, conditional on enforcement, whether to divert the fine should the offender be detected. In our design, no sanction is applied for enforcers' diverting of the fine's proceeds. At the time of deciding whether to take, the potential offender is informed about *both* of the enforcer's decisions.

Treatment CORRUPTION differs from FLAT and REWARD in that the enforcer makes decisions not only about investing but also about diverting the fine meant to benefit society. Accordingly, in this treatment, the enforcer's receipt of the fine results from the enforcer's choice, where diverting presumably is a norm infraction. Clearly, it may matter for

[9] The arrangement is standard for a large number of experiments. For example, subjects concerned solely about monetary payoffs should not transfer money in dictator games (e.g., Engel, 2011), abstain from costly punishment in one-shot interactions (e.g., Friehe and Utikal, 2018) or contribute positive amounts to public goods (e.g., Fehr and Gächter, 2000).

[10] Several potentially imperfect ways of representing a redistribution to society at large in the laboratory are possible. For example, redistributing fine revenue to other participants in the same session could influence subjects' earnings expectations and increase the relevance of beliefs about takings by other session subjects, meaning a loss of control for the experimenter. We opted for the charity to avoid such complications. Most important, in criminal proceedings in Germany, for example, judges can, under certain circumstances, actually decide that fines will be transferred to a charity (see, e.g., Weigend, 2001). Les Resto du Coeur is a very well-known French charitable association that mainly provides food to the poor and homeless.

the potential offenders' choice whether the regime is designed such that enforcers always receive the fine revenue (as in treatment REWARD) or whether enforcers can profit only from their own misconduct (as in treatment CORRUPTION).

Table 1 provides an overview of our treatments.

To sum up, the game proceeds as follows:

- Part 1: a real-effort task is completed;
- Parts 2–5: subjects participate in taking games; subjects play one treatment in Parts 2–3 and another treatment in Parts 4–5 with, first, a direct-response format and, second, a strategy-method format;
- Part 6: information about participants is collected.

3.3 Further information about treatment-independent parts 1 and 6

3.3.1 Part 1: The real-effort task

We include the real-effort task, in which participants work for their endowment of 500 points, to create a notion of entitlement (e.g., Faillo et al., 2019; Falk & Fischbacher, 2002; Oxoby & Spraggon, 2008). We selected a threshold task to create symmetric endowments (e.g., Duersch & Müller, 2015). The task consists of correctly counting the number of zeros in tables consisting of 100 randomly ordered zeros and ones. Each participant must report the correct number in at least two tables within 10 minutes. Failing to do so leads to the subject's exclusion from the experiment. As Abeler et al. (2011) emphasize, that boring task does not require any prior knowledge, performance is easily measurable, comes at a positive cost of effort, and there is little learning possibility. Circumstantially, all participants were successful in the task.

3.3.2 Part 6: Subject characteristics

In Part 6, participants make choices in incentivized tasks without strategic payoff interdependence and provide us information about their attitudes on different aspects related to their decisions in the taking game. Specifically, we inquire about their morality (using the items from Traxler & Winter, 2012) and justice attitudes (using the different points of view conceptualized by Schmitt et al., 2010). Regarding morality, we present eight different norm violations taken from Traxler and Winter (2012) to our participants (drunk driving, hazardous waste disposal, speeding, steeling newspapers, absenteeism from work, evading TV fees, fare dodging, and tax evasion), and ask for their evaluation of the behaviors on a scale from 1 (very severe) to 10 (not severe). In our empirical analysis, we rely on the inverted sum of the eight responses as our morality measure (with 80 as the possible maximum). With regard to justice sensitivity, we follow Schmitt et al (2010), and ask for participants' assessment of eight situations that are framed to entail some level of unfairness. Participants should rate their agreement with statements such as "*It makes me angry when others are undeservingly better off than m*e" and "*I feel guilty when I am better off than others for no reason*" using a six-point rating scale ranging from 0 (not at all) to 5 (exactly). We rely on the eight-item sum as our justice sensitivity measure (with 40 as the possible maximum).

In addition, we incentivize responses to a social value orientation (SVO) test (Murphy et al., 2011) and a risk elicitation task (Eckel & Grossman, 2008). Finally, we collected

Table 2 Groups and subjects per treatment sequence

Treatment Sequence	Flat–Reward	Reward–Flat	Corruption–Flat	Reward–Corruption	Corruption–Reward
Sessions	6	6	8	3	3
Groups	36	33	46	17	17
Subjects	108	99	138	51	51

observations on subjects' demographics. Summary statistics on the subjects' characteristics are shown in Tables A.1 and A.2. in Appendix A.

3.4 Procedures

We conducted the experiment in the laboratory of the University Panthéon-Sorbonne, Paris, between May and October 2019.[11] Participants were administered and recruited via ORSEE (Greiner, 2015) from the laboratory's subject pool. We collected data from 447 participants in 26 sessions, 298 of which were active participants (i.e., either potential offenders or enforcers). Subjects' mean age was about 35 years. Our participants (about 60% female) primarily were students from various fields of study but also included a few retirees. Table 2 presents the number of groups and subjects per treatment sequence. Overall, we had 115 groups for FLAT, 103 groups for REWARD, and 80 groups for CORRUPTION (with each group going through two treatments).

A typical session lasted around 80 minutes (including payment). During the experiment, subjects earned points that were converted to euros at a conversion rate of one euro cent per point. Participants were paid in cash at the end of each session, the average earning being around 11.75 euro. The experimenter made the charity donation after the relevant sessions were concluded, and participants were informed at the beginning of each session that the charity donation would be made in accordance with the result of the experiment.

4 Hypotheses

We now briefly formulate hypotheses for our design. We start with the benchmark of "standard" agents before we introduce behavioral aspects that might lead to different conjectures regarding how the deterrent effect depends on who receives fine revenues. We generally assume that risk-neutral agents' payoffs depend on expected income and moral disutility.

4.1 Hypotheses relying on standard assumptions

For our standard agents, moral concerns stem only from the agent's own behavior. Potential offenders incur moral disutility if they take points from the victim. Following Fees et al. (2018), we assume moral disutility m_j^O for an offender j, where m_j^O is distributed according

[11] The experimental software was developed at the Laboratoire d'Economie Expériementale de Paris (LEEP) by development engineer Maxim Frolov using.NET/c# and php/WAMP technologies.

to the cumulative distribution function $G(m_j^O)$ on $[0, \overline{m}^O]$. We assume that law enforcers perceive some moral duty to fulfill their tasks by creating a positive detection probability. That assumption accords with the experimental findings of Engel and Zhurakhovska (2017). Thus, enforcer i bears moral disutility m_i^E if she does not create a positive detection probability, where m_i^E is distributed according to the cumulative distribution function $H(m_i^E)$ on the support $[0, \overline{m}^E]$.

Using backward induction, we first consider the potential offender's decision for the two cases in which the enforcer did or did not invest in establishing a positive detection probability. If the detection probability is positive, the potential offender takes the amount b from the victim if

$$y + b - pF - m_j^O > y, \tag{1}$$

where y is initial income, p the detection probability, and F the fine in the event of detection.[12] Taking occurs if moral disutility is less than the expected monetary gain ($m_j^O < b - pF$), which results in a taking rate equal to $G(b - pF)$. If the detection probability is zero, the potential offender takes from the victim if

$$y + b - m_j^O > y, \tag{2}$$

which results in a taking rate equal to $G(b)$. The deterrent effect from the enforcer's investment is given by the difference in taking rates and equal to

$$G(b) - G(b - pF). \tag{3}$$

In other words, enforcement reduces the probability that a potential offender has a moral disutility level small enough to make taking preferable by the difference in relation (3). The difference is independent of how the fine's proceeds are allocated. We summarize:

Hypothesis 1 ***(Standard theory)*****:** (a) Enforcement's deterrent effect on the potential violators' taking choice is the same in treatments FLAT and REWARD. (b) Enforcement's deterrent effect on the potential violators' taking choice is independent of whether the enforcer diverts the fine in treatment CORRUPTION.

Next, we examine the enforcer's choice in Stage 1. That choice depends on who receives fine's revenue, and we concentrate on treatments FLAT and REWARD. If fine revenue benefits society at large (treatment FLAT), the enforcer creates a positive detection probability if

$$y - c > y - m_i^E, \tag{4}$$

where c denotes the enforcer's investment cost. The enforcer invests if the monetary cost falls short of moral disutility from non-enforcement ($c < m_i^E$), and the share of enforcers

[12] In this section, we consider norm violation as a binary choice. In our experiment, potential offenders selected the amount, if any, to take, knowing that the detection probability and the level of the sanction are independent of the amount taken (if positive). In other words, any taking should occur at the maximum level when moral disutility from taking is fixed or at most linear in the level of taking.

investing in a positive detection probability amounts to $1 - H(c)$. In contrast, if enforcers receive fine revenues (treatment REWARD), enforcement takes place if

$$y - c + qpF > y - m_i^E, \tag{5}$$

where q is the enforcers' belief about the taking rate.[13] Accordingly, enforcers invest if the moral disutility from non-enforcement exceeds the expected net enforcement cost, $m_i^E > c - qpF$. That critical level implies that the share of investing enforcers amounts to $1 - H(c - qpF)$. Comparing the two regimes, we find that the probability that a potential enforcer has a moral disutility level large enough to trigger the investment rises if financial incentives are provided. The difference amounts to

$$H(c) - H(c - qpF) > 0. \tag{6}$$

In addition, we note that the lower critical level for moral disutility in treatment REWARD implies that, conditional on investing in the detection probability, the enforcers' average moral concern is lesser.

We summarize:

Hypothesis 2 ***(Enforcer behavior):*** (a) More enforcers invest in a positive detection probability in treatment REWARD than in treatment FLAT. (b) Among the enforcers who invest in a positive detection probability, morality scores are lower in treatment REWARD than in treatment FLAT.

4.2 Taking behavioral effects into account

For "standard" potential offenders, we find that deterrence is unaffected by the identity of fine recipients. Below, we outline how the consideration of two behavioral ideas can fuel the expectation that enforcement's deterrent effect should be affected by the recipients' identity even when the expected sanction remains the same. First, we consider the possibility that potential offenders respond to different normative meanings of enforcement in the two remuneration regimes.[14] Second, we allow for the possibility that potential offenders suffer disutility from unfavorable income comparisons along the lines of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).

Regarding the normative meaning of enforcers' investment decisions, note that people are willing to forfeit resources to prevent or punish non-compliance with norms about which they care (e.g., Fehr & Fischbacher, 2004). Against that background, enforcers who invest in treatment FLAT signal that they perceive "not taking" to be a strong social norm because creating a positive detection probability is costly and not associated with any personal material benefits. In contrast, subjects who enforce in treatment REWARD may be driven by a profit-seeking motive and thus are, on average, less concerned about social norms than subjects who enforce in treatment FLAT. The

[13] With rational expectations and perfect information, we have $q = G(b - pF)$.

[14] Related contributions in the realm of normative meaning are varied and include work by Bénabou and Tirole (2016) on the relationship between beliefs and action; Adriani & Sondereggger (2018), in which parents' behavior signals the normative content of human action to their children; and the stream of literature on leading-by-example (on leaders influencing the normative meaning of behavior, see, e.g., Dannenberg 2015).

potential offender may consider the enforcer's belief about the strength of the norm to be informative about social conventions more generally. The receiver of fine revenue then will affect the normative information that can be deduced from enforcement and thereby enforcement's deterrent effect. If the potential offender cares about aligning own choices with strong social norms (as seems to be true in many circumstances; see, e.g., Kimbrough & Vostroknutov, 2016), then taking after enforcers invest in detection imposes larger norm non-compliance costs in treatment FLAT, which should result in a stronger deterrent effect.

To be specific, assume that the offender's moral costs, m_j^O, are augmented by an amount s according to the normative signal obtained from the enforcer's investment decision. Consequently, given enforcement, the offender takes from the victim if

$$y + b - pF - \left(m_j^O + s\right) > y. \tag{7}$$

This ranking implies a taking rate amounting to $G(b - pF - s)$. If the normative meaning of enforcement is higher in FLAT than in REWARD (i.e., if $s^{FLAT} > s^{REWARD}$), as argued above, we obtain that the deterrent effect of the enforcer's enforcement on the potential violators' taking choice is greater in FLAT than in REWARD:

$$G(b) - G\left(b - pF - s^{\text{FLAT}}\right) > G(b) - G\left(b - pF - s^{\text{REWARD}}\right) \tag{8}$$

In treatment CORRUPTION, the enforcer decides whether to divert the fine, thereby providing a strong signal about the normative meaning of enforcement. Once again, if potential offenders care about how the enforcer perceives the norm, the deterrent effect of enforcement will be affected. In summary, the foregoing considerations suggest:

Hypothesis 3 ***(Normative meaning of enforcement):*** (a) Enforcement's deterrent effect on the potential violators' taking choice is greater in treatment FLAT than in treatment REWARD. (b) In treatment CORRUPTION, enforcement's deterrent effect on the potential violators' taking choice is greater when the enforcer does not divert the fine.

Next, consider the possibility that potential offenders experience disutility from unfavorable income comparisons. In that case, we expect enforcement's deterrent effect to be greater in treatment REWARD than in FLAT. The intuition runs as follows: consider an offender who takes the maximum possible amount such that $F = b > c$, and the victim's loss is equal to $2b$ in our experiment. Taking creates advantageous inequity for the offender vis-à-vis the victim and, in the case of no detection, vis-à-vis the enforcer in both treatments. In contrast, in the event of detection, the offender enjoys a higher income vis-à-vis the enforcer only in treatment FLAT (enforcer income: $-c$; offender income:$y + b - F = y$) but not in treatment REWARD (enforcer income: $y - c + F$; offender income:y). Denoting with β the measure of how an unfavorable income comparison translates into disutility, we find that the offender takes with a positive detection probability in treatment REWARD if

$$y + b - pF - m_j^O - p\beta(F - c) > y, \tag{9}$$

where $y + b - pF$ is expected income from taking and $\beta(F - c)$ represents the weighted difference in enforcer and offender payoffs in the case of detection which occurs with probability p. Inequality (9) results in a taking rate equal to $G(b - pF - p\beta(F - c))$,

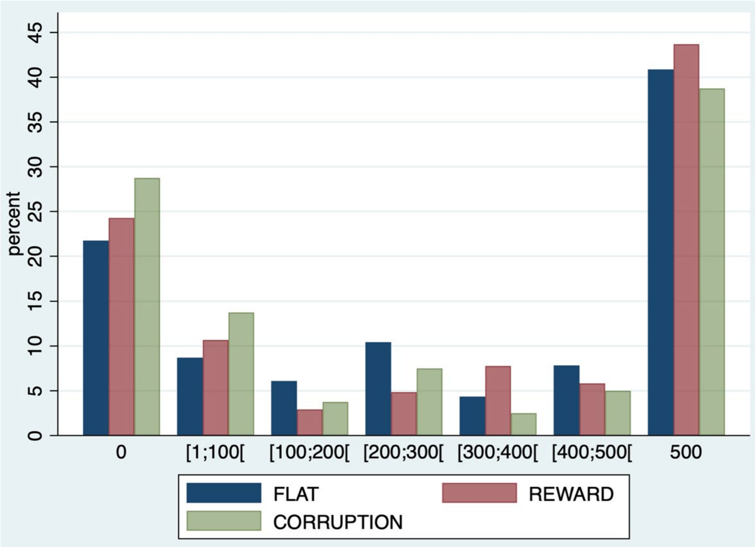


Fig. 1 Distribution of amounts taken by treatment

which compares to a taking rate of $G(b - pF)$ in treatment FLAT. The deterrent effect of a positive detection probability is stronger in treatment REWARD than in treatment FLAT according to

$$G(b) - G(b - pF - p\beta(F - c)) > G(b) - G(b - pF). \tag{10}$$

A similar logic applies to the treatment CORRUPTION when the enforcer decides to divert the fine after detecting a crime.

Hypothesis 4 ***(Inequity aversion):*** (a) Enforcement's deterrent effect on the potential violators' taking choice is greater in treatment REWARD than in treatment FLAT. (b) In treatment CORRUPTION, enforcement's deterrent effect on the potential violators' taking choice is greater when the enforcer diverts the fine.

Thus, inequity aversion leads to predictions that are opposite to those stemming from a theory building on a normative meaning of enforcement and a desire to comply with norms. The mechanisms are quite different as well. The normative meaning of enforcement implies that potential violators behave differently in treatments REWARD and FLAT because the enforcer's decision to create the positive detection probability conveys a stronger signal about the norm in FLAT than in REWARD. In contrast, the theory of inequity aversion assumes that agents anticipate and seek to avoid disadvantageous payoff comparisons. In our setup, enforcement can induce disadvantageous inequity for the offender only in treatment REWARD, such that enforcement would create more deterrence in REWARD than FLAT.

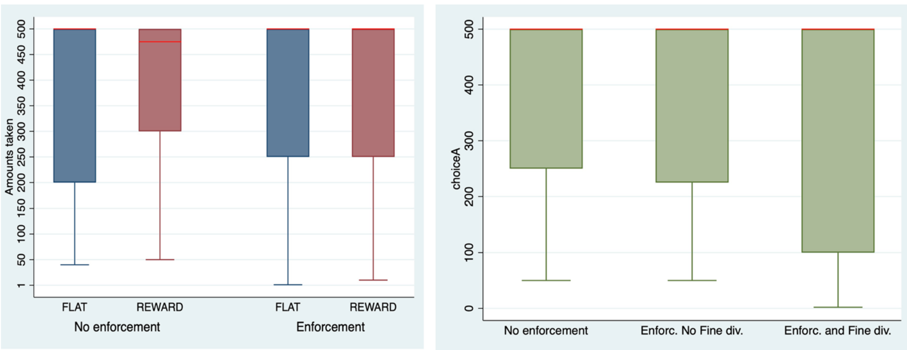
In summary, we note that the standard framework in Sect. 4.1 predicts a null result for the effect of the enforcer's remuneration on offenders' taking for a given detection probability. With the incorporation of selected behavioral aspects, it is possible to obtain different effect signs. All in all, the direct link between enforcement's deterrent effect and the enforcers' remuneration is unclear, making an empirical analysis particularly valuable.

5 Results

In our analysis, we focus on observations from Parts 2 and 4, that is, the parts adopting the direct response method.[15] Below, we enter a dummy variable that indicates whether the observation is collected from Part 2 or 4, whose coefficient is not significant in any empirical model.

5.1 A glance at the data

Figures 1 and 2 summarize offenders' taking behavior in our experiment. As is clear from Fig. 1, which shows the amounts taken by treatment across different choices by the enforcer, a very large share of subjects takes the maximum amount (i.e., 500



a FLAT and REWARD **b CORRUPTION**

Fig. 2 Amounts taken conditional on taking by treatment and enforcer's choice

[15] We emphasize the strategy method parts of our experiment because we initially were concerned about data that were too unbalanced, a concern that proved to be unwarranted. Building on Brandts and Charness (2011), differences between the direct-response method and the strategy method can be expected because emotions are involved, subjects make only a few conditional choices in the real world, and the game's structure is one-shot. However, we find no significant differences between direct response and strategy methods in our experiment for the potential violators. See Parts A and D in our Supplementary Material for a thorough analysis.

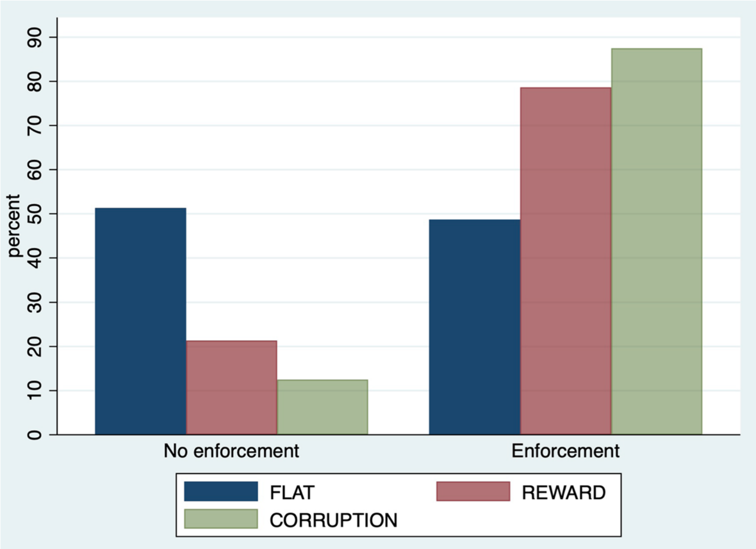


Fig. 3 Enforcement choices by treatment

points) from the victim, and another large share chooses not to take at all. On average, only about 34% choose some amount of taking between those extremes across treatments, whereas 24.5% choose not to take at all and 41.28% take the maximum amount. Figure 2 reports amounts taken conditional on taking and enforcers' choice, confirming that the majority of violators opt for the maximum taking amount. Indeed, boxplots show that the median taking always is 500 except in treatment REWARD when the enforcer does not invest in enforcement (where the median is 475 points). From those distributions, we deduce that a dummy variable that is equal to one if the offender takes and zero otherwise is not associated with a substantial loss of information because the choice seems to reduce to not taking or taking the maximum amount for most individuals. Thus, when we analyze offenders' taking below, we will focus on results from probit regressions.

In Fig. 3, we illustrate the distribution of enforcers' choices. As expected, monetary incentives increase the likelihood of enforcement. In treatment CORRUPTION, enforcers had to decide about enforcement and, conditional on enforcement, whether to divert the fine into their own pockets. In our experiment, 65% of enforcers who invested in detection chose to divert the fine to their own benefit.

Table A.3 in our Supplementary Material provides detailed statistics about the offenders' taking and the enforcers' choices in our experiment.

5.2 Empirical analysis

5.2.1 Deterrent effect of enforcement

We now test Hypotheses 1, 3, and 4 on individual choice data and subjects' characteristics elicited in Part 6 of the experiment. Our experimental design and the distribution

Table 3 Coefficients and AMEs from random-effects probit regressions for taking for FLAT & REWARD treatments

	(1) Coef	(2) AME	(3) Coef	(4) AME
REWARD	−0.103	−0.007	2.276	0.003
	(0.458)	(0.030)	(1.409)	(0.032)
Enforcement	−1.916***	−0.078**	−1.123	−0.151***
	(0.546)	(0.034)	(0.798)	(0.043)
REWARD x enforcement			−2.946*	
			(1.708)	
SVO angle			−5.096***	−0.366***
			(1.951)	(0.127)
Morality			−0.034	−0.002
			(0.033)	(0.002)
Justice sensitivity			0.029	0.002
			(0.045)	(0.003)
Risk attitude			0.249	0.018
			(0.195)	(0.011)
Age			0.093***	0.007***
			(0.035)	(0.002)
Female			−1.599*	−0.111*
			(0.940)	(0.061)
Order			−0.435	−0.031
			(0.554)	(0.036)
Constant	4.845***		3.390	
	(0.456)		(2.978)	
Observations	218	218	218	218
Number of id	149	149	149	149

Coefficients and AMEs from random-effects probit regressions. The dependent variable is equal to one (zero) if (no) taking occurred. REWARD is a dummy variable equal to one if the observation is from treatment REWARD. Enforcement is a dummy variable equal to one if the enforcer created a positive detection probability;

*, **, *** indicate significance at the 10%, 5%, and 1% levels. Cluster robust standard errors are shown in parentheses

of amounts taken (see, e.g., Fig. 1) suggest using a random-effects probit model because the observation of a given subject across several decision points creates a panel dataset.

Entering $Taking_{it}$ as a dummy variable equal to one if subject i decided to take points from the victim in period t, we specify

$$\begin{aligned}\text{Taking}^*_{it} =& \alpha + \beta_1 \text{Treatment}_{it} + \beta_2 \text{Enforcement}_{it} + \beta_3 \text{Treatment}_{it} * \text{Enforcement}_{it} \\ &+ \beta_4 \text{Order}_{it} + \gamma Z_i + u_i + \varepsilon_{it}\end{aligned} \tag{11}$$

with $Taking^*_{it}$ as the latent variable and Z_i as a vector containing subject-specific characteristics such as *SVO angle, Morality, Justice Sensitivity, Risk Attitude, Age,* and *Sex*. Remember that a higher SVO angle indicates greater prosociality (Murphy et al., 2011) and that the individual's risk attitude was elicited using the lottery selection task by Eckel

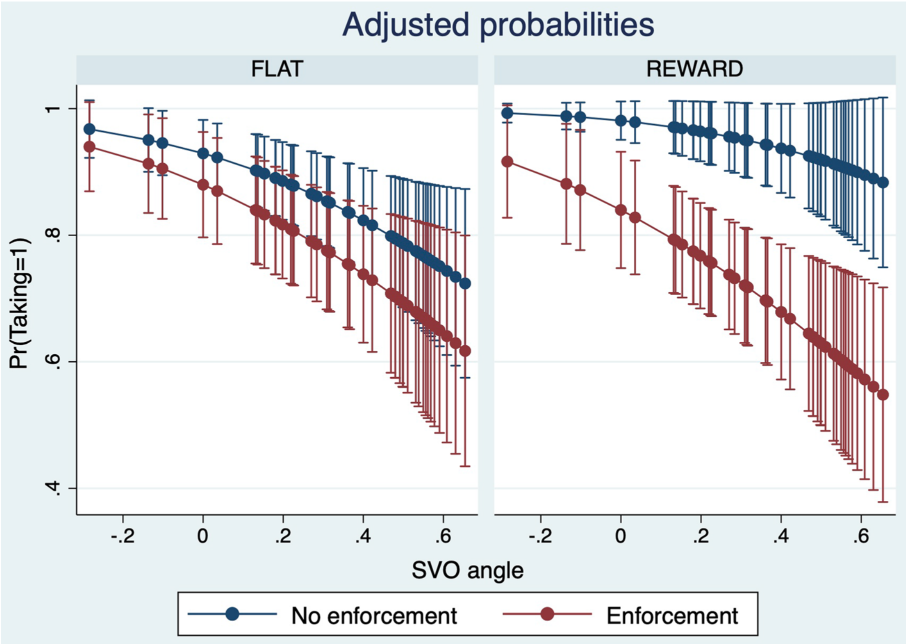


Fig. 4 Adjusted taking probabilities by SVO angle (95% confidence intervals)

and Grossman (2008) where a larger number represents greater risk tolerance. For each potential violator i, $Treatment_{it}$ varies across periods, and $Enforcement_{it}$ may vary as well. To control for order effects, $Order_{it}$ is entered as a dummy variable that is equal to one if the potential offender's decision stems from Part 4 of the experiment (instead of Part 2). The equation contains two error terms, the conventional error term ε_{it} and the individual-specific unobservable effect u_i. The relationship between the latent variable $Taking^*_{it}$ and the observed binary variable $Taking_{it}$ is:

$$\text{Taking}_{it} = \begin{cases} 1 \text{ if Taking}^*_{it} > 0 \\ 0 \text{ if Taking}^*_{it} \leq 0 \end{cases} \tag{12}$$

First, we compare treatments FLAT and REWARD. In Table 3, we report the estimated coefficients and average marginal effects (AMEs) of our variables of interest. The latter describes the average difference in the expected probability of taking associated with a one-unit change in the variable of interest, averaged over the distribution of the other variables in the dataset. We find that enforcement reduces the taking probability while the enforcers' remuneration scheme has no significant direct effect. However, our research question concerns whether the deterrent effect of enforcement depends on enforcer remuneration. Column (3) in Table 3 reports a significant and negative coefficient for the interaction between enforcement and the REWARD treatment.

However, according to Norton et al. (2004), the statistical significance of the interaction effect cannot be tested with a simple t-test on the coefficient of the interaction term in non-linear models. For that purpose, Williams (2012) suggests visual display and analysis of differences between treatments. Along those lines, Fig. 4 plots the adjusted probabilities for both treatments by enforcer's choice and SVO angle with 95% confidence intervals. Except at very small SVO angles, the confidence intervals are not overlapping in treatment REWARD, whereas they do so in treatment FLAT. As another piece of evidence, we report a significant interaction between Enforcement and REWARD from random-effects linear probability models in our Supplementary Material (Table B.1). Thus, we conclude that enforcement reduces the probability of taking in treatment REWARD but not in treatment FLAT.

The interaction described above is evident at the extensive margin but is insignificant at the intensive margin. That conclusion follows from regressions that enter the amounts taken as the dependent variable (for details, see Table B.2 in our Supplementary Material). Moreover, the effect seems to be weaker in our strategy-method experiment such that the interaction effect is insignificant (see Table A.4 in our Supplementary Material), which seems consistent with differences in outcomes reported for other experimental settings (e.g., Brandts & Charness, 2011).

We summarize:

Result 1: Enforcement's deterrent effect is (weakly) larger in treatment REWARD than in treatment FLAT.

A larger deterrent effect of enforcement in REWARD is inconsistent with Hypotheses 1 and 3, which consider the standard agent and the possibility that norm enforcement conveys normative meaning to the potential offender. However, Result 1 is consistent with Hypothesis 4, according to which potential offenders are inequity averse. Our finding is consistent with the idea that agents respond to income comparisons and that they behave in order to avoid situations perceived as personally disadvantageous. To the extent that treatment REWARD ultimately can result in an unfavorable income comparison from the potential offender's perspective, the deterrent effect of enforcement is stronger than in treatment FLAT.

Regarding personal characteristics, we find that the SVO angle is relevant for explaining variation in the binary taking variable in our empirical model, whereas the coefficients for Justice Sensitivity, Risk Attitude, and Morality are insignificant. The finding that Morality does not contribute to explaining taking behavior is surprising (but we will show that the Morality score is relevant for the enforcement decision). We admit that the finding may result from our imprecise measure of the complex construct "morality".

Next, we address treatment CORRUPTION. As that treatment differs structurally from treatments FLAT and REWARD, we restrict our analysis to treatment CORRUPTION and focus on gauging how enforcement's deterrent effect depends on the enforcer's diversion choice (i.e., the endogenous remuneration change).

The model we estimated for that purpose is

$$\text{Taking}_i^* = \alpha + \beta_1 \text{Enforcer behavior}_i + \beta_2 \text{Order}_{it} + \gamma Z_i + u_i + \varepsilon_i, \tag{13}$$

where enforcer behavior is a nominal categorical variable that is either *No Enforcement*, *Enforcement & No Fine Diversion* or *Enforcement & Fine Diversion*. Potential offenders participate in treatment CORRUPTION at most once, so we estimate a standard probit

Table 4 Coefficients and AMEs from probit regressions for taking (treatment CORRUPTION)

	(1) Coef	(2) AME	(3) Coef	(4) AME
No enforcement	−0.379	−0.089	−0.097	−0.019
	(0.601)	(0.148)	(0.652)	(0.127)
Enforcement & fine diversion	−0.877**	−0.254**	−0.857	−0.213**
	(0.432)	(0.100)	(0.531)	(0.107)
Justice sensitivity			−0.063**	−0.016***
			(0.027)	(0.006)
SVO angle			−1.376**	−0.360**
			(0.617)	(0.155)
Morality			0.021	0.006
			(0.014)	(0.004)
Risk attitude			−0.123*	−0.032*
			(0.071)	(0.018)
Age			0.003	0.001
			(0.013)	(0.003)
Female			0.371	0.098
			(0.380)	(0.101)
Order			0.538	0.134
			(0.528)	(0.121)
Constant	1.221***		2.036*	
	(0.394)		(1.189)	
Observations	80	80	80	80

Coefficients and AMEs from probit regressions. The dependent variable is equal to one (zero) if (no) taking occurred. Enforcement & fine diversion is dummy variable equal to one if the enforcer created a positive detection probability and diverts the fine. No Enforcement is dummy variable equal to one if the enforcer did not create a positive detection probability;

*, **, *** indicate significance at the 10%, 5%, and 1% levels. Cluster robust standard errors are shown in parentheses

model. In Table 4, the reference level for enforcer behavior is *Enforcement & No Fine Diversion*.[16]

To test our hypotheses, we evaluate the AME associated with the enforcer's behavior. Enforcers' fine diversions raise enforcement's deterrent effect significantly: we report a significant AME for *Enforcement & Fine Diversion* in Columns (2) and (4) (which is the difference in the deterrence effect versus *Enforcement & No Fine Diversion*). Results from linear probability models reported in our Supplementary Material corroborate that finding.

[16] In Table A.5, we show that the results are comparable when information from the strategy method also is considered.

Table 5 Coefficients and AMEs from random-effect probit regressions for enforcement (treatments FLAT, REWARD and CORRUPTION)

	(1) Coef	(2) AME	(3) Coef	(4) AME	(5) Coef	(6) AME
REWARD	0.926***	0.296***	0.919***	0.296***	3.130***	0.298***
	(0.217)	(0.061)	(0.219)	(0.063)	(1.091)	(0.062)
CORRUPTION	1.332***	0.386***	1.326***	0.386***	4.412***	0.388***
	(0.256)	(0.058)	(0.264)	(0.062)	(1.282)	(0.062)
Morality			−0.018**	−0.005**	0.001	−0.005**
			(0.008)	(0.002)	(0.010)	(0.002)
REWARD x morality					−0.036**	
					(0.017)	
CORRUPTION x morality					−0.049**	
					(0.020)	
Justice sensitivity			−0.001	−0.000	−0.001	−0.000
			(0.012)	(0.003)	(0.012)	(0.003)
SVO angle			0.101	0.028	0.104	0.028
			(0.361)	(0.099)	(0.365)	(0.098)
Risk attitude			0.055	0.015	0.056	0.015
			(0.043)	(0.012)	(0.043)	(0.011)
Age			0.001	0.000	0.001	0.000
			(0.007)	(0.002)	(0.007)	(0.002)
Female			0.022	0.006	0.028	0.008
			(0.190)	(0.052)	(0.195)	(0.052)
Order			−0.052	−0.014	−0.021	−0.006
			(0.181)	(0.050)	(0.183)	(0.049)
Constant	−0.031		0.630		−0.494	
	(0.132)		(0.703)		(0.823)	
Observations	298	298	298	298	298	298
Number of id	149	149	149	149	149	149

The dependent variable is binary and equal to one when the enforcer opted for investment in detection. REWARD and CORRUPTION are dummy variables indicating treatments;

*, **, *** indicate significance at the 10%, 5%, and 1% level. Cluster robust standard errors are shown in parentheses

Result 2: I*n treatment CORRUPTION, enforcement's deterrent effect is larger when the enforcer opts for fine diversion.*

That finding is inconsistent with Hypotheses 1 and 3 (i.e., the assumption of a standard agent or the conveyance of normative meaning by enforcement) but consistent with Hypothesis 4 (i.e., potential offenders' inequity aversion).

5.2.2 Enforcers' investment

Next, we consider enforcer behavior and test Hypothesis 2. In treatments FLAT and REWARD, enforcers choose between no investment and investment in different periods.

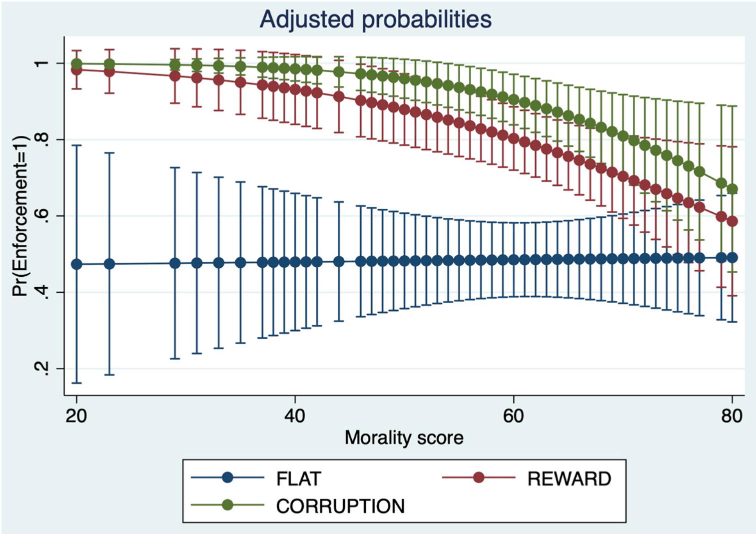


Fig. 5 Adjusted enforcement probabilities by *Morality* score (95% confidence intervals)

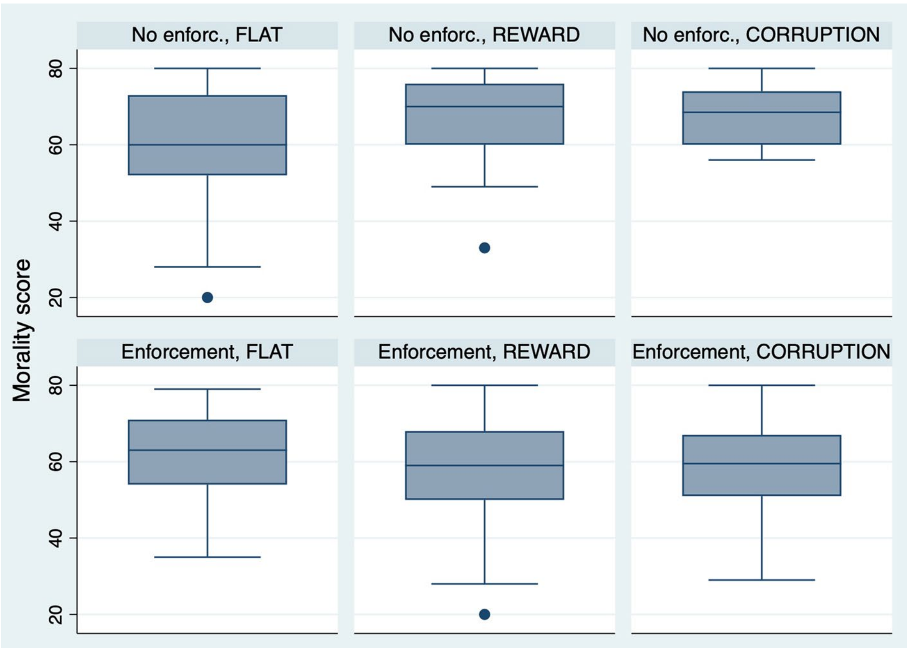


Fig. 6 Enforcers' morality scores by treatment and enforcement choice

In our analysis of enforcement choices, we are particularly interested in which types of individuals willingly incur private costs to enforce the norm. To classify individuals, we rely on our *Morality* score since it informs us about subjects' perceptions regarding norm violations in different domains and has proven very meaningful in other research (Traxler & Winter, 2012).

For our analysis, we consider a random-effects probit model of the following form:

$$\text{Enforcement}^*_{it} = \alpha + \beta_1 \text{Treatment}_{it} + \beta_2 \text{Treatment}_{it} * \text{Morality}_i + \beta_3 \text{Order}_{it} + \gamma Z_i + u_i + \varepsilon_{it} \quad (14)$$

with $Enforcement^*_{it}$ as the latent variable and Z_i as a vector containing subject-specific characteristics.

The results reported in Table 5 are set against treatment FLAT as the base category. Table 5 confirms the intuition that enforcement is more likely in treatments that provide personal financial motives. In Columns (3) and (4), we also find that a higher *Morality* score significantly reduces the probability of enforcement. We aim at resolving that puzzle in Column (5), where we show that the negative effect stems from the treatments with financial incentives.

To interpret the interaction between the treatment and *Morality* score, we plot the adjusted probabilities of enforcement by *Morality* score with 95% confidence intervals (Fig. 5). Whereas the enforcement probability is significantly higher in the treatments with financial incentives at low and intermediate morality scores, that no longer is true at high morality scores.

Result 3: *(a) More enforcers invest in the detection probability in treatments REWARD and CORRUPTION than in treatment FLAT. (b) The morality scores of enforcers who invest in the detection probability in treatments offering financial rewards (REWARD and CORRUPTION) is less than that of enforcers in treatment FLAT.*

Result 3 (a) is confirmed in random-effects linear probability models reported in our Supplementary Material and consistent with our Hypothesis 2. The higher enforcement effort in treatments with financial enforcement incentives accords with previous evidence (e.g., Harvey, 2020). Moreover, as Result 3 (b) suggests, the remuneration scheme influences the characteristics of active enforcers (we focus on the enforcer's *Morality* score). Figure 6 confirms Result 3 (b). That figure shows the boxplots of enforcers' morality scores by treatment and enforcement choice. The boxplots indicate that subjects with low morality scores are more prevalent among enforcers who chose to invest in a positive detection probability in the treatments with financial incentives (REWARD and CORRUPTION) than in treatment FLAT. Those observations align with our regression results in Table 5 and are consistent with the positive self-selection results presented in Friebel et al. (2019) for police applicants in Germany, where no monetary rewards for enforcement successes are granted.

6 Discussion and concluding remarks

This paper relies on experimental data to explore whether enforcement's *deterrent effect* depends on the enforcer remuneration regime even when the expected sanction is constant across regimes. The previous literature focused on the implications of varying enforcers' financial incentives on the level of enforcement effort and the allocation of enforcement resources. We compare enforcement's deterrent effect for a given expected sanction when the enforcer receives the fine to the case in which the fine is transferred to society at large while the enforcer receives a lump sum payment.

Our paper presents some evidence that questions the standard assumption in the theory of law enforcement that the identity of the recipient of fine revenue is irrelevant for compliance choices if the levels of the detection probability and the fine are held constant. The deterrent effect of enforcement seems to be stronger if enforcers receive fine revenue. With respect to possible mechanisms, our results seem consistent with potential offenders' inequity aversion. In addition, we confirm that regimes offering financial incentives to enforcers lead to more enforcement effort but also a different mix of subjects choosing to engage in enforcement. With financial incentives, subjects with low morality scores are more likely to invest in enforcement.

Our experiment identifies the consequences of the enforcer remuneration choice in a laboratory experiment. Clearly, we must be careful when deriving implications for the real world from our findings. Enforcement's deterrent effect seems to be greater in treatment REWARD, but we do not view that result as a basis for recommending financial incentives for enforcers. In fact, some evidence exists that subjects are more likely to take in REWARD in the absence of enforcement effort. More important, our design allows only two possible levels for the detection probability (zero or 50%), so that we do not know how the differential effects reported herein will bear out at other levels of the detection probability. In reality, potential offenders may not be informed about actual enforcement choices and likely will expect more vigorous enforcement in the presence of personal financial incentives, an effect that we purposefully exclude from our design. It is also important to acknowledge that our design highlights a possible selection effect at the stage where individuals consider becoming a law enforcer. One could fear that by implementing financial incentives, agents with lower moral values would be selected as law enforcers. We also must be careful in extrapolating our results because of possible issues of external validity that often arise in laboratory experiments also apply here (such as greater scrutiny; see, e.g., Levitt & List, 2007).

Moreover, in order to concentrate on the pure remuneration effect, we omitted the possibility that enforcers extort compliant agents on purpose. In reality, such concerns may arise. The theoretical literature and the findings of Xiao (2013) make clear that the possibility of Type II legal errors (i.e., being fined despite compliance) weakens deterrence. Therefore, incentivizing law enforcers financially might entail a series of consequences that our experiment can capture only partially. We believe that it may be interesting for future work to investigate the implications of such additional issues in isolation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11127-022-01000-5.

Acknowledgements We gratefully acknowledge the very helpful comments and suggestions received from two anonymous reviewers, the handling editor (Jordan Adamson), and the editor-in-chief (William F. Shughart II). In addition, we thank the participants of the Experimental Law & Economics Workshop at George Mason University, in particular Gary Charness, Gregory DeAngelo, Johanna Mollerstrom, and Murat Mungan.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review, 101*(2), 470–492.

Adamson, J., & Rentschler, L. (2021). How policing incentives affect crime, measurement, and justice. Working Paper, https://ssrn.com/abstract=3628595.

Adriani, F., & Sonderegger, S. (2018). Signaling about norms: Socialization under strategic uncertainty. *Scandinavian Journal of Economics, 120*, 685–716.

Baicker, K., & Jacobsen, M. (2007). Finders keepers: Forfeiture laws, policing incentives, and local budgets. *Journal of Public Economics, 91*, 2113–2136.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*, 169–217.

Becker, G. S., & Stigler, G. J. (1974). Law enforcement, malfeasance and compensation of enforcers. *Journal of Legal Studies, 3*, 1–18.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives, 30*, 141–164.

Benson, B. L., & Rasmussen, D. W. (1995). Police bureaucracies, their incentives, and the war on drugs. *Public Choice, 83*, 21–45.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*, 166–193.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics, 14*, 375–398.

Carpenter, D. M., Knepper, L., Erickson, A. C., & McDonald, J. (2015). *Policing for profit: The abuse of civil asset forfeiture* (2nd ed.). Institute for Justice.

Dannenberg, A. (2015). Leading by example versus leading by words in voluntary contribution experiments. *Social Choice and Welfare, 44*, 71–85.

Dharmapala, D., Garoupa, N., & McAdams, R. H. (2016). Punitive police? Agency costs, law enforcement, and criminal procedure. *Journal of Legal Studies, 45*, 105–141.

Dickinson, D. L., Masclet, D., & Villeval, M. C. (2015). Norm enforcement in social dilemmas: An experiment with police commissioners. *Journal of Public Economics, 126*, 74–85.

Duersch, P., & Müller, J. (2015). Taking punishment into your own hands: An experiment. *Journal of Economic Psychology, 46*, 1–11.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization, 68*, 1–17.

Engel, C. (2011). Dictator games: A mega study. *Experimental Economics, 14*, 583–610.

Engel, C., & Zhurakhovska, L. (2017). You are in charge: Experimentally testing the motivating power of holding a judicial office. *Journal of Legal Studies, 46*, 1–50.

Faillo, M., Rizzolli, M., & Tontrup, S. (2019). Thou shalt not steal: Taking aversion with legal property claims. *Journal of Economic Psychology, 71*, 88–101.

Falk, A., & Fischbacher, U. (2002). "Crime" in the lab—detecting social interaction. *European Economic Review, 46*, 859–869.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980–994.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868.

Friebel, G., Kosfeld, M., & Thielmann, G. (2019). Trust the police? Self-selection of motivated agents into the German Police Force. *American Economic Journal: Microeconomics, 11*, 59–78.

Friehe, T., & Mungan, M. C. (2021). The political economy of enforcer liability for wrongful police stops. *Journal of Public Economic Theory, 23*, 141–157.

Friehe, T., & Utikal, V. (2018). Intentions undercover—Hiding intentions is considered unfair. *Journal of Behavioral and Experimental Economics, 73*, 11–21.

Garoupa, N. (1997). A note on private enforcement and type-I error. *International Review of Law and Economics, 17*, 423–429.

Garoupa, N., & Klerman, D. (2002). Optimal law enforcement with a rent-seeking government. *American Law and Economics Review, 4*, 116–140.

Garret, T. A., & Wagner, G. A. (2009). Red ink in the rearview mirror: Local fiscal conditions and the issuance of traffic tickets. *Journal of Law and Economics, 52*, 71–90.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*, 114–125.

Harvey, A. (2020). Fiscal incentives in law enforcement. *American Law and Economics Review, 22*, 173–210.

Helland, E., & Tabarrok, A. (2004). The fugitive: Evidence on public versus private law enforcement from bail jumping. *Journal of Law and Economics, 47*, 93–122.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association, 14*, 608–638.

Koyama, M. (2014). The law and economics of private prosecutions in industrial revolution England. *Public Choice, 159*, 277–298.

Landes, W. M., & Posner, R. A. (1975). The private enforcement of law. *Journal of Legal Studies, 4*, 1–46.

Langlais, E., & Obidzinski, M. (2017). Law enforcement with a democratic government. *American Law and Economics Review, 19*, 162–201.

Leibbrandt, A., & Lopez-Perez, R. (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior and Organization, 84*, 753–766.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives, 21*, 153–174.

Makowsky, M. D., & Stratmann, T. (2009). Political economy at any speed: What determines traffic citations? *American Economic Review, 99*, 509–527.

Makowsky, M. D., & Stratmann, T. (2011). More tickets, fewer accidents: How cash-strapped towns make for safer roads. *Journal of Law and Economics, 54*, 863–888.

Makowsky, M. D., Stratmann, T., & Tabarrok, A. (2019). To serve and collect: The fiscal and racial determinants of law enforcement. *Journal of Legal Studies, 48*, 189–216.

Mast, B. D., Benson, B. L., & Rasmussen, D. W. (2000). Entrepreneurial police and drug enforcement policy. *Public Choice, 100*, 285–308.

Miceli, T. J., & Johnson, D. (2016). Asset forfeiture as a law enforcement tool. *Contemporary Economic Policy, 34*, 119–126.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*, 771–781.

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal, 4*, 154–167.

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior and Organization, 65*, 703–713.

Parrillo, N. R. (2013). *Against the profit motive: The salary revolution in American Government, 1780–1940*. Yale University Press.

Polinsky, A. M., & Shavell, S. (2007). The theory of public enforcement of law. In A. M. Polinsky & S. Shavell (Eds.), *Handbook of Law and Economics 1*. North Holland.

Polinsky, A. M. (1980). Private versus public enforcement of fines. *Journal of Legal Studies, 9*, 105–127.

Rizzolli, M., & Stanca, L. (2012). Judicial errors and crime deterrence: Theory and experimental evidence. *Journal of Law and Economics, 55*, 311–338.

Romaniuc, R., Farrow, K., Ibanez, L., & Marciano, A. (2016). The perils of government enforcement. *Public Choice, 166*, 161–182.

Schildberg-Hörisch, H., & Strassmair, C. (2012). An experimental test of the deterrence hypothesis. *Journal of Law, Economics, and Organization, 28*, 447–459.

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The justice sensitivity inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211–238.

Traxler, C., & Winter, J. (2012). Survey evidence on conditional norm enforcement. *European Journal of Political Economy, 28*, 390–398.

US Department of Justice (2015). Investigation of the Ferguson Police Department.

Weigend, T. (2001). Sentencing and punishment in Germany. In M. Tonry & R. S. Frase (Eds.), *Sentencing and Sanctions in Western Countries.* Oxford University Press.

Wickelgren, A. L. (2003). Justifying imprisonment: On the optimality of excessively costly punishment. *American Law and Economics Review, 5*, 377–411.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal, 12*, 308–331.

Xiao, E. (2013). Profit-seeking punishment corrupts norm obedience. *Games and Economic Behavior, 77*, 321–344.

Yahagi, K. (2021). Law enforcement with motivated agents. *International Review of Law and Economics, 66*, 105982.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.