Journal of Behavioral and Experimental Economics 77 (2018) 107–121

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Hard labor in the lab: Deterrence, non-monetary sanctions, and severe procedures

Matteo Rizzolli[*,a], James Tremewan[b]

[a] LUMSA University, Rome, Italy
[b] University of Auckland, Vienna University of Economics and Business, Austria

ARTICLE INFO

JEL classification:
D01
K14
K40
K41
K42
Non-monetary sanctions
Real effort task
Hard labor
Punishment
Adjudicative procedures
Expressive law

ABSTRACT

We experimentally investigate two questions that must be understood to effectively implement important normative prescriptions of optimal deterrence theory: i) does a non-monetary punishment and a fine of equivalent monetary value produce the same level of deterrence, and ii) should severe procedures, which maximize correct convictions of guilty defendants, be preferred to lenient procedures, which minimize errors in cases against innocent defendants? We examine these questions in an experiment where potential thieves face the possibility of punishment. As a non-monetary sanction, we require convicted individuals to perform a tedious real effort task. In the monetary treatments, sanctions are instead fines, which are based on individuals' willingness to pay to avoid the real effort task to ensure comparability with the non-monetary treatment. The second manipulation in our experiment concerns the balance of errors in the adjudicative procedure (i.e. the conviction of innocents and acquittal of guilty individuals). We find that stealing is reduced most effectively by a sanction regime that combines non-monetary sanctions with a severe procedure. Our data are consistent with the notion that both monetary punishment and pro-defendant sanction regimes are less effective in communicating moral condemnation of an act.

1. Introduction

Becker's (1968)) seminal work initiated an important stream of literature on optimal deterrence (Polinsky and Shavell, 2009; Garoupa, 2014) that has informed reforms in criminal laws and criminal procedures in many countries, in particular in the US. Since the 1970s, major reforms aimed at increasing deterrence have involved the reintroduction of the death penalty in many US states and mandatory minimum sentences, as well as 'tough on crime' and 'three-strikes' laws. As a result, the number of prisoners has grown from fewer than 200,000 in 1972 to more than 1.6 million in 2010, which, accounting for general population growth, implies a four-fold increase in the incarceration rate (Travis et al., 2014). The theory of optimal deterrence relies on some strong rationality assumptions and has included behavioral economics only recently (Garoupa, 2003; Dhami and al Nowaihi, 2013). Attempts at testing the deterrence hypothesis abound (Chalfin and McCrary, 2017), but some aspects of the theory are inherently difficult to falsify (Levitt and Miles, 2007). Scholars have only recently begun to test the theory with rigorous experimental designs (see the literature review below), and this paper continues in their footsteps by further testing two prominent policy-relevant aspects of the theory.

The theory of optimal deterrence prescribes policymakers to substitute non-monetary sanctions (N-MSs) with monetary sanctions (MSs) whenever possible. The intuition is simple: if we substitute an N-MS with an MS while keeping deterrence constant, we reduce social costs because MSs are transfers from convicted individuals to the rest of society, while N-MSs are costly for both the convicted and society. Beyond prison, other forms of N-MSs include home detention (under probation or parole), mandated community service, and mandatory drug treatment. The use of shaming sanctions have also been explored, such as publishing convicted individuals' names on billboards, in newspapers, or even on broadcast television (Kahan, 1996; 2005-2006; Kahan and Posner, 1999). Since the 1970s, seemingly in line with the prescriptions of Becker's theory, the actual use of MSs has unquestionably increased in all western countries' criminal laws and procedures (Beckett and Harris, 2011), while in the US, there has also been a parallel increase in N-MSs (Harris et al., 2010; Ruback, 2015).

The degree to which MSs can substitute for N-MSs is, therefore, an important empirical question that has very relevant policy implications. From a traditional economic perspective, replacing an N-MS with an MS equal to a potential criminal's willingness to pay to avoid the sanction should provide the same level of deterrence. However, this is unclear

* Corresponding author.
E-mail addresses: m.rizzolli@lumsa.it (M. Rizzolli), james.tremewan@auckland.ac.nz (J. Tremewan).

https://doi.org/10.1016/j.socec.2018.09.011
Received 30 August 2017; Received in revised form 18 September 2018; Accepted 24 September 2018
Available online 29 September 2018
2214-8043/ © 2018 Elsevier Inc. All rights reserved.

behaviorally because, for example, an N-MS may send a stronger normative message than an MS, which could be regarded simply as a 'cost of doing business.' With our first treatment manipulation, we aim to experimentally test whether a non-monetary punishment and a fine of equivalent monetary value produce the same level of deterrence by comparing subjects' propensity to steal in a laboratory setting under the different sanction types.

The second aspect of the theory of optimal deterrence tested in this paper concerns the impact of the two possible types of judicial errors: conviction of innocents and acquittal of guilty individuals. We refer to pro-defendant adjudicative procedures that keep the ratio of innocents convicted to guilty persons acquitted low as *lenient*, and those that place greater weight on convicting the guilty as *severe*.[1] Legal scholars have been discussing this trade-off for a long time, consistently supporting lenient procedures because wrongful convictions are typically regarded as worse mistakes than wrongful acquittals from a moral perspective (Volokh, 1997; Nicita and Rizzolli, 2014).

However, are wrongful convictions also worse than wrongful acquittals in terms of lost deterrence? The standard model of optimal deterrence (see Section 3) shows that both types of errors are detrimental to deterrence. The intuition behind this result is that while increasing the probability of wrongful acquittal increases the expected payoff from committing a crime, increasing the probability of wrongful conviction reduces the expected payoff of abiding by the law. Png (1986) showed that, under some simplifying assumptions, the marginal effects of the two types of errors on deterrence are equally negative. Thus, if deterrence is the primary policy concern, there is no particular reason to favor lenient procedures over severe ones. Taking the deterrence argument to the letter, several scholars have questioned some of the pillars of current pro-defendant procedures and suggested reducing the standard of proof (Ognedal, 2005) and removing other pro-defendant procedural mechanisms.

The hypothesized equivalence of the two types of judicial errors is difficult to test empirically because it is nearly impossible to disentangle correct from wrongful convictions in real world procedures.[2] Therefore, the impact of wrongful convictions on general deterrence cannot be credibly assessed with naturally occurring data. However, this can be done in the laboratory, where guilt and innocence is known to the experimenter, and error probabilities can be exogenously manipulated.

Our experiment takes the following form: we first elicit subjects' willingness to pay (*wtp*) to avoid a tedious real effort task and then, after the decision to steal is made and the adjudication process is completed, impose either an MS equal to their *wtp* or an N-MS in the form of an obligation to carry out the same real effort task ('hard labor') on those determined guilty. Hence, we can compare N-MSs with MSs that have the same disutility, a comparison that would be impossible outside the laboratory. We also exogenously manipulate the conviction probabilities for both innocent and guilty individuals to determine whether the nature of the adjudicative procedure has an asymmetric impact on deterrence.

Our results suggest that an N-MS and an equivalent MS do not produce the same level of deterrence: an N-MS coupled with a severe error structure reduces crime more than an N-MS with a lenient error structure or either of the MS procedures. All sanction regimes appear to increase the bimodality of amounts stolen by reducing the proportion of subjects stealing but increasing the amount stolen among those who do. The increase in the amount stolen, conditional on stealing, is greater with N-MSs coupled with a severe error structure, leading to very little difference between treatments in the total amounts stolen. The data suggest that treatment differences are not driven by risk-aversion but are consistent with the idea that severe procedures and N-MSs communicate a greater degree of social condemnation. In line with this, we find no evidence that the marginal effect of increasing the level of sanctions on reducing the propensity to steal differs between treatments as one would expect if the reduction in crime was driven by the sort of cost-benefit analysis assumed by standard deterrence theory.

2. Literature review

This paper touches upon two main streams of literature, theoretical and experimental, which we address in turn. The first stream concerns the law and economics approach to optimal deterrence. Here we focus on two aspects of the theory that are relevant to our experiment: i) the nature of the sanction (i.e. MS versus N-MS) and ii) the nature of the procedure (i.e. lenient versus severe).

The claim made on purely utilitarian grounds that MSs should substitute for N-MSs as much as possible dates to Becker (1968). The same general argument was stated formally by Friedman (1981) and Polinsky and Shavell (1984a): from the social viewpoint, MSs are preferred to N-MSs because the former are transfers from the sanctioned individuals to society, while the latter are costly to both the individual and society. Note that, from the individual's viewpoint, they are equivalent since they both decrease utility.[3] While the deterrence theory of justice favors MS over N-MS, retributivist theories of punishment justify the use of N-MSs, claiming punishment should fit the crime, not only in size (the principle of proportionality) but also in kind (Avio, 1993; Posner, 1980; Wittman, 1974). Furthermore, incapacitation theories of punishment justify N-MSs, such as prison and banishment, as they are effective means to avoid further damage to society (Ehrlich, 1981; Kan, 1996; Miceli, 2010; Mungan, 2012; Shavell, 2015).

Expressive theories of law (Cooter, 1998; Nance, 1997) may also justify the use of N-MSs. According to McAdams (2000): "law changes behavior by signaling the underlying attitudes of a community or society," and therefore, N-MSs such as prison may send a stronger message of condemnation than MSs, which can appear to be "no more than a luxury tax on the prohibited activity" (Markel and Flanders, 2010). The idea that the introduction of a monetary sanction can crowd out moral considerations is supported by the findings of Gneezy and Rustichini (2000) and the subsequent literature on incentives (positive and negative) and motivation crowding-out (Bowles and Polania-Reyes, 2012; Sugden, 2018; Grant, 2018).

The second aspect of the theory of optimal deterrence that is relevant to our experiment concerns the nature of the adjudicative procedure. Png (1986) first noted the symmetric effect of two types of errors on deterrence: while wrongful acquittals increase the returns of engaging in crime, wrongful convictions decrease the returns of staying honest. This result has been included in all major surveys on optimal deterrence (See Garoupa, 1997; Polinsky and Shavell, 2009).[4] This simple result has a straightforward policy implication: a lenient procedure that produces 'too many' wrongful acquittals to contain wrongful convictions is no more deterring than a symmetric severe

[1] We borrow the lenient versus severe terminology from the performance rating literature (Murphy and Balzer, 1989; Prendergast, 1999).

[2] For a review of the literature see Rizzolli (2018).

[3] Specific exceptions exist to this general rule: N-MSs are preferred over MSs when the defendants are either too poor to pay the fine, and thus, cannot be deterred (Polinsky and Shavell, 1984b; Shavell, 1986), or they are too rich to be deterred, as well as when defendants can hide their wealth, or their wealth depends on their human capital (Levitt, 1997b), when the authority needs to signal its commitment to a sanctioning strategy (D'Antoni and Galbiati, 2007), and in presence of corruption (Garoupa and Klerman, 2004). Fines are also often considered *unjust* because the rich are able to pay fines, whereas the poor serve jail sentences (Levitt, 1997a). All these exceptions justify the use of N-MSs within the deterrence framework.

[4] There exists an ongoing debate on whether this prediction (i.e. that both errors are equally detrimental to deterrence) withstands closer theoretical scrutiny. The debate originated with Lando (2006) and was followed by Garoupa and Rizzolli (2013) and Lando and Mungan (2017).

procedure that tolerates 'too-many' wrongful convictions to limit wrongful acquittals. However, actual criminal procedures in modern democratic countries are overwhelmingly lenient and pro-defendant, and therefore, some extensions have been proposed to reconcile the gulf between theory and evidence. These extensions encompass the inclusion of risk-aversion (Rizzolli and Stanca, 2012), loss-aversion (Nicita and Rizzolli, 2014), cost of sanctions (Rizzolli and Saraceno, 2013), and the chilling of desirable behavior (Mungan, 2011). Regarding the view of punishment as conveying social condemnation, a lenient procedure weakens the message by signaling that the state is not overly concerned about the offender's misconduct (Markel and Flanders, 2010), whereas a severe procedure erring on the side of convictions signals that an action is sufficiently unacceptable that innocents can be sacrificed to ensure the punishment of norm violators.

To summarize, the deterrence theory of justice prescribes the use of MSs to reduce social costs under certain conditions. However, other well established theories of justice might be evoked to explain the persistent use of N-MSs. The deterrence theory also prescribes minimizing overall errors with no exclusive preoccupation with wrongful acquittals.

The second stream of literature this paper touches upon is the experimental one. The experimental literature on punishment is vast, but relatively few papers are closely related to ours. Regarding the type of sanction, Masclet et al. (2003) show that an individual expression of disapproval can partly substitute monetary sanctions in a voluntary contribution mechanism (VCM) game. Their result has been supported by several papers (Noussair and Tucker, 2005; Peeters and Vorsatz, 2013; Pérez and Kiss, 2012); however, what they call an N-MS is more accurately described as a form of second-party shaming punishment. Fiedler and Haruvy (2017) implement third-party punishment in an investment game and show that monitoring/shaming alone induces much of the compliance obtained with standard punishment mechanisms. Cinyabuguma et al. (2005) also introduce an N-MS in a VCM game in the form of banishment. In our project, MSs and N-MSs are both imposed by an anonymous central authority so that shaming cannot play a role.

Several experimental papers test the deterrence hypothesis in the lab using the same inverse dictator game we use as a baseline. However, all these papers only deal with MSs. The only study we are aware of that compares MSs and N-MSs is Montag and Tremewan (2018), which investigates whether people are willing to condition the level of punishment on the subjective experience of the convicted.

Regarding the effects of procedures on deterrence, the experimental literature is quite limited. Grechenig et al. (2010) first show with a lab experiment that judicial errors greatly undermine deterrence in a VCM-type game. Rizzolli and Stanca (2012) disentangle the effects and find that wrongful convictions are more detrimental to deterrence than wrongful acquittals, but they do not reject the hypothesis that risk-aversion alone could explain this asymmetry. Marchegiani et al. (2016) find the same effect in a principal-agent setting. In contrast, Markussen et al. (2016) find the two types of errors have a symmetric effect in a VCM framework. An interesting finding from Markussen et al. (2016) is that when subjects vote for which type of error they will face, wrongful convictions become even less detrimental to deterrence, while wrongful acquittals become more so. Although this effect could be due to selection, it is also consistent with the idea that errors signal social norms and that these signals are given more legitimacy when chosen democratically.

3. Model and predictions

Our experiment compares the level of stealing in four different treatments by manipulating two key variables: the type of sanction and the error structure of the adjudicative procedure. For clarity, we refer to sanction *types* (i.e. MS versus N-MS), sanction *procedure* (i.e. lenient versus severe), and sanction *regime* (i.e. the type/procedure combination).[5] In the following paragraphs we present the standard model of optimal deterrence then discuss behavioral reasons why one may expect our results to differ from the resulting predictions.

Let b be the gains from crime and s_{nm} the N-MS, while s_m is the MS. In the model, the two differ in as much as the disutility of the N-MS is additively separable from the utility of the monetary gain from crime, which is not the case for the MS. This kind of separability between monetary and non-monetary elements in a utility function is standard in economics, for example, in principal-agent models where utility from income and disutility from effort are additive. Let w be the level of wealth at the time of the decision to commit the crime. Let ε_1 be the probability of a wrongful conviction and ε_2 be the probability of a wrongful acquittal. $A = 1 - \varepsilon_1 - \varepsilon_2$ is a measure of the accuracy of the adjudicative process. Note that two very different states of the world, with either a *lenient* or a *severe* procedure, can be characterized by the same level of accuracy if

$$A^{LENIENT} = 1 - \varepsilon_1^{LOW} - \varepsilon_2^{HIGH} = 1 - \varepsilon_1^{HIGH} - \varepsilon_2^{LOW} = A^{SEVERE}.$$

We assume individuals to be utility maximizers who decide whether to commit the crime or abstain on purely self-regarding grounds.

Monetary sanctions. We consider individuals with standard utility functions à-la von Neumann-Morgenstern who, if convicted, must pay a monetary sanction s_m. Each agent weighs his own returns from committing the crime ($EU_g = \varepsilon_2 U(w+b) + (1-\varepsilon_2)U(w+b-s_m)$) against the expected returns of abstaining from crime ($EU_i = (1-\varepsilon_1)U(w) + \varepsilon_1 U(w-s_m)$). Deterrence is achieved if $EU_i \geq EU_g$, and thus, if

$$\varepsilon_1[U(w) - U(w-s_m)] + \varepsilon_2[U(w+b) - U(w+b-s_m)] \leq U(w) - U(w+b-s_m).$$

The left-hand side is increasing in both ε_1 and ε_2, demonstrating that both types of errors jeopardize deterrence. However, is there any difference in the relative impact of the two errors? The 'weights' of the two errors, $U(w) - U(w-s_m)$ and $U(w+b) - U(w+b-s_m)$, respectively, are projections onto the vertical axis of different tracts of the utility function, the latter higher on the curve than the former. If agents are risk-neutral, the utility function in wealth is linear ($EU(.)'' = 0$), the two weights are the same, and therefore the two errors have the same impact on deterrence. Risk-averse agents, however, have a concave utility function ($EU(.)'' < 0$), which implies that $U(w) - U(w-s_m) > U(w+b) - U(w+b-s_m)$, and therefore ε_1 has a larger detrimental impact on deterrence than ε_2. Conversely, risk-loving agents ($EU(.)'' > 0$) are more sensitive to changes in ε_2 than ε_1. Thus, for risk-averse agents facing procedures of equal accuracy, a severe procedure is less deterring than a lenient procedure; for risk-neutral agents, severe and lenient procedures are equally deterring; and for risk-loving agents, a severe procedure is more deterring than a lenient procedure (see Nicita and Rizzolli (2014) and Rizzolli (2018) for detailed discussions).

Non-monetary sanctions. When the sanction is not monetary, the results are very similar to those under risk neutrality once we assume separability in the monetary and non-monetary elements in the utility function. The utility of the action choices available, remaining law-abiding or committing crime, are $EU_i = (1-\varepsilon_1)U(w) + \varepsilon_1[U(w) - s_{nm}]$ and $EU_g = \varepsilon_2 U(w+b) + (1-\varepsilon_2)[U(w+b) - s_{nm}]$, respectively. Deterrence is obtained for $U(w+b) - U(w) \leq (1-\varepsilon_1-\varepsilon_2)s_{nm}$. Note that this inequality is identical to that which determines deterrence under monetary sanctions if agents are risk-neutral and $s_m = s_{nm}$.

While a simple *homo economicus* model predicts that there should be no difference in stealing behavior between treatments, previous work has suggested several reasons why these predictions could fail (see

[5] When discussing experimental results, the terms regime and treatment will be used interchangeably.

Section 2). First, non-monetary sanctions may send a stronger message of social condemnation than fines, making them more effective in reducing stealing. Regarding adjudicative procedures, risk aversion among subjects would reduce the effectiveness of severe as opposed to lenient procedures in the monetary treatment; however, if severe procedures indicate greater social condemnation of stealing, then these will generate greater deterrence.

4. Experimental design

The experiment consisted of four parts: i) a slider task; ii) a procedure designed to elicit subjects' willingness to pay (*wtp*) to avoid repeating the slider task; iii) three decisions about how much to steal from another subject; and iv) an elicitation of risk preferences. These incentivized tasks were followed by an unincentivized questionnaire. No feedback was given until after all tasks and the questionnaire were completed to minimize the possibility of outcomes affecting responses in later parts of the experiment. Subjects who had to repeat the slider task due to the *wtp* elicitation, the realized stealing decision, or both were required to do so before collecting their final payment.[6] No subject refused to repeat the tasks when required to do so.

Slider task. The first part of the experiment involved the 'slider task' (see Gill and Prowse, 2012 for further details). Subjects were paid 5 Euros to place 96 onscreen sliders precisely in the middle of their respective lines. No time limit was imposed.

Willingness to pay elicitation. In this part of the experiment, we implemented a Becker-DeGroot-Marschak mechanism to elicit subjects' evaluation of the effort task. Subjects were given an extra 6 Euros, which they were told they could use for this task. They made 13 binary decisions between placing a further 96 sliders or paying a sum of money that varied between 0 and 6 Euros in increments of 50 cents. One of the thirteen decisions was chosen at random at the end of the experiment, and subjects either kept the full 6 Euros and repeated the slider task or kept only the remainder of the money after the relevant sum was deducted. Choices were forced to be 'consistent' in the sense that if a certain sum of money was chosen to be paid in preference to repeating the slider task, then all lesser sums would also have to be chosen.

Stealing decisions. Subjects were faced with three decisions concerning how much to steal from another randomly selected subject. They were informed that they could take up to 5 Euros from another subject then asked "Do you want to take from the other participant? If yes, how much do you want to take?" In the first decision of all subjects (*No − sanction*), there was no possibility of punishment. As all treatments were identical up to and including this decision, it gave us a reliable control for social and moral concerns related to stealing that were unrelated to the fear of punishment. The second decision is where we introduced our treatment manipulations. We implemented a two-by-two design, varying the type of sanctions in the stealing decisions (MS and N-MS) between subjects and the error structure of the procedure (*Severe* and *Lenient*) within subjects. The order in which the different error structures were introduced was counterbalanced within each session to control for possible order effects.

Subjects were informed that after they made the decision, their choice would be audited, and if the audit failed, possibly as the result of a mistake, they would be punished. In the *Severe* treatment, they would be punished with 90% probability if they stole and 50% probability if they did not. In the *Lenient* treatment, these probabilities were 50% and 10%, respectively.[7] The third decision was made under the alternative error structure.

In the N-MS treatment, the punishment was to place an additional 96 sliders. In the MS treatment, the sanction was based on the individual's *wtp* to avoid repeating the task that was elicited in the previous phase. Our *wtp* protocol elicited an interval of € 0.50 width within which the true *wtp* should lie, and the fine was set at the midpoint of this interval (i.e. the average of the maximum sum the individual was willing to pay to avoid repeating the slider task and the minimum sum where they preferred to repeat the task rather than pay). Subjects were not informed at any time that the fine was based on their earlier decisions.[8]

At the end of the experiment, half of the subjects were randomly selected to have one of their three taking choices implemented, which was also randomly selected, while the other half were the possible victims of the theft. It was made clear in advance that at most one decision was going to be implemented, so each of the three decisions should have been treated independently.

Risk preference elicitation. This part of the experiment consisted of eight binary choices between two lotteries. For each choice, subjects who chose lottery A (i.e. the safe lottery) won either € 1.80 or € 2.25 with equal probability; those who chose lottery B (i.e. the risky lottery) won either € 0.92 or a sum that increased from € 2.74 to € 5.96 across the choices, again with equal probability. At the end of the experiment, one of these decisions was randomly selected, and the chosen lottery played out.[9] Choices were forced to be consistent such that if a subject chose lottery B when the larger sum they could win was a given amount, they had also to choose lottery B when that sum was even greater.

Controls and questionnaire. Finally, subjects had to complete a questionnaire containing the Cognitive Reflection Test (Frederick, 2005) to measure individuals' specific cognitive ability, a selection of questions from the Domain-Specific Risk-Taking (DOSPERT) measure (Blais and Weber, 2006), two versions of the trolley dilemma (Edmonds et al., 2014), and some standard demographic controls.

Procedures. The experiment was programmed in z-Tree (Fischbacher, 2007) and run at the Vienna Center for Experimental Economics between April and October 2015. Subjects were invited using ORSEE (Greiner, 2015). A total of 70 subjects participated in the MS treatment, and 76 participated in the N-MS treatment. The ages of our subjects ranged between 20 and 43 years (26 on average), and 43% were female. Sessions lasted approximately one hour, and subjects earned between € 6.25 and € 25.

Table 1
Proportion of subjects who steal and average amounts stolen. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ indicates the difference from the decision under no sanctions (exact unconditional Z test for proportion stealing and Wilcoxon signed-rank tests for amounts stolen).

	Steal	Amount stolen	Amount stolen|Steal
M-S No Sanction	0.657	2.67	4.07
M-S Lenient	0.571	2.64	4.61***
M-S Severe	0.557*	2.49	4.46***
N-MS No Sanction	0.750	3.11	4.14
N-MS Lenient	0.632**	3.01	4.76***
N-MS Severe	0.487***	2.43**	5.00***

[6] Instructions and screenshots can be found in appendix Appendix B.

[7] The choice of these error probabilities strikes a balance between keeping accuracy constant while creating enough treatment variation between the two procedures and maintaining a positive level of deterrence.

[8] If subjects stated that they were willing to pay every amount, then all we know is that their *wtp* is greater than € 6. We used the same formula for these subjects as the others and assumed their *wtp* was € 6.25.

[9] We chose this method, based on Drichoutis and Lusk (2012), as it has been shown to give the most reliable results for list-based risk elicitation methods (Csermely and Rabas, 2016).

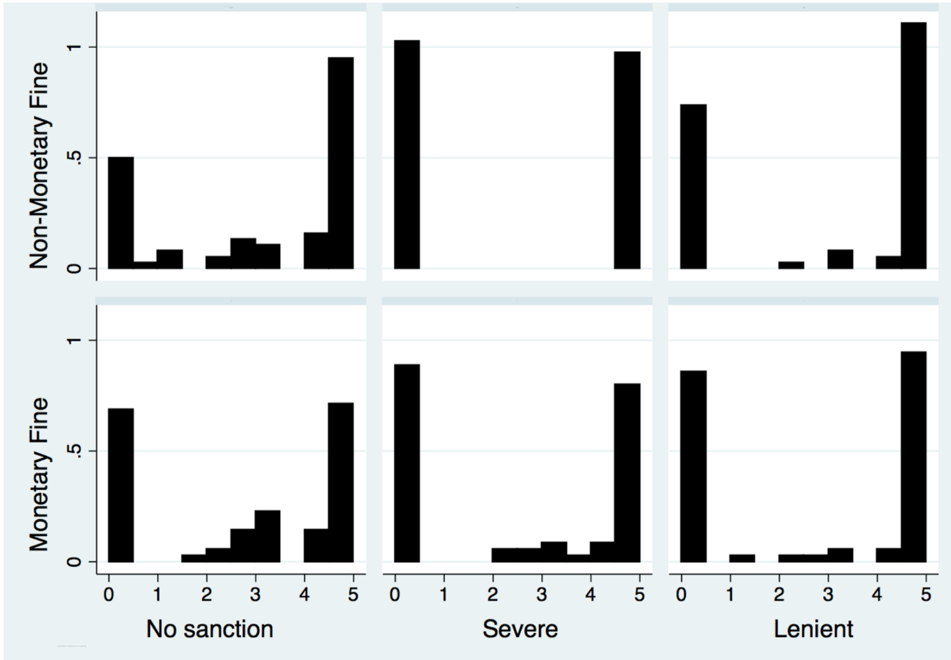


Fig. 1. Distributions of amount stolen under the three procedures and two sanction types.

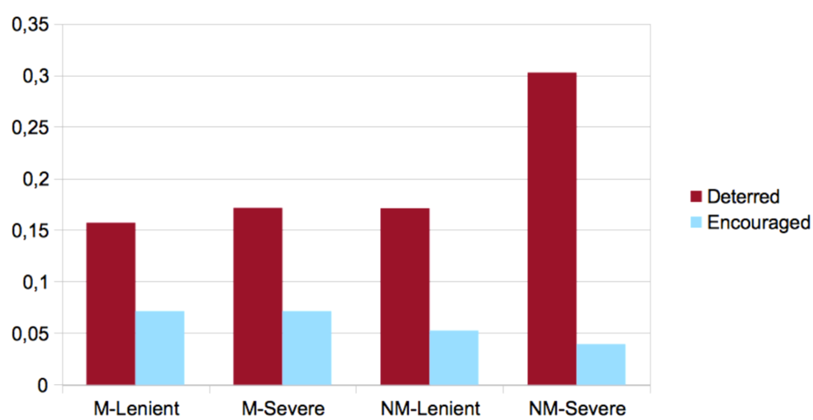


Fig. 2. Proportion of subjects stealing in each treatment.

5. Results

We begin by giving an overview of the data and testing the impact of introducing sanctions. We then proceed to our main questions of interest in comparing the efficacy of the four sanction regimes in reducing the propensity to steal and the amounts stolen. Finally, we analyze the possible mechanisms underlying our treatment effects.

5.1. Overview of data

We find no evidence of order effects for any of the four treatments in the proportion of subjects stealing or amount stolen, so we have pooled the data. Details of the statistical tests can be found in Appendix A.1. For simple treatment comparisons of the deterrence level to be valid, it is important that the distributions of sanctions, based on the *wtp* of subjects to avoid repeating the slider task, are similar across the MS and N-MS sessions. The average *wtp* was € 1.66, and the distribution did not differ significantly between treatments (MS: 1.72; N-MS: 1.60; WMW test: $p = 0.472$).[10] The average number of ‘safe choices’ was 3.9 (risk neutral subjects should choose three). The distribution of safe choices differed significantly across treatments, with subjects in MS displaying more risk aversion (MS: 4.5; N-MS: 3.4; WMW test: $p < 0.01$).[11] The distribution of correct answers to the Cognitive Reflection Test did not differ between treatments (MS: 1.46; N-MS: 1.62; WMW test: $p = 0.371$).

Table 1 shows the proportion of subjects stealing, the average amount stolen, and the average amount stolen conditional on stealing for each of the three decisions, separated by sanction type. We first note that there is an apparent but not statistically significant difference between levels of stealing without sanctions between the MS and N-MS

[10] In the *wtp* elicitation, a small number of subjects chose to place additional

(*footnote continued*)
sliders rather than pay nothing, decisions most likely made in error. We think it probable that these subjects had a genuinely low *wtp* and made an error only on the first decision, so we assume that their *wtp* is between 0 and € 0.5. All results are robust to dropping these subjects.

[11] Risk aversion was elicited after stealing decisions, so it is possible that exposure to the different sanction types induced different responses to the risk-aversion task. Controlling for risk aversion in our regression analysis fails to explain our treatment effects.

Table 2
Regression analysis of the probability of stealing (Probits - marginal effects)

Variables	(1) *Steal*	(2) *Steal*	(3) *Steal*	(4) *Steal*
MS*severe*	-0.0143	-0.0181	-0.0155	-0.0484
	(0.0557)	(0.0708)	(0.0747)	(0.0841)
N-MS*lenient*	0.0610	0.0139	0.00491	0.0260
	(0.0818)	(0.0896)	(0.0909)	(0.141)
N-MS*severe*	-0.0843	-0.160*	-0.177*	-0.144
	(0.0829)	(0.0899)	(0.0905)	(0.143)
Stole (no sanction)		0.527***	0.544***	0.545***
		(0.0647)	(0.0636)	(0.0627)
Sanction size			-0.0680***	-0.0641
			(0.0262)	(0.0447)
Sanction size x MS*severe*				0.0159
				(0.0541)
Sanction size x N-MS*lenient*				-0.0110
				(0.0583)
Sanction size x N-MS*severe*				-0.0181
				(0.0603)
Observations	292	292	292	292
Treatment comparisons				
MS*S* vs N-MS*L*	0.83	0.13	0.05	0.26
MS*S* vs N-MS*S*	0.72	2.44	3.10*	0.42
N-MS*L* vs N-MS*S*	6.75***	6.70***	6.83***	2.98*

Notes: stole (no sanction) = 1 if subject stole in decision without sanctions. *Treatment Comparisons* are chi-square statistics from tests of equality of coefficients of the treatment dummies. Standard errors clustered by subject in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

treatments (66% and 75%, respectively; Z test[12] $p = 0.253$). We remind the reader that both treatments were identical up to and including this decision, so any difference would be due to random variation rather than the treatment manipulation; however, it does suggest the importance of controlling for this variable when making comparisons between sanction types.

The proportion of subjects stealing is reduced by all four sanction regimes, although the difference is not significant for MS *Lenient*, and only weakly significant for MS *Severe*. However, a sign test fails to reject the equality of medians in all cases, so the rejection of the Wilcoxon sign-ranked test may not have resulted from a change in central tendency; instead, it was other changes in the shape of the distributions. Interestingly, while reducing the probability of stealing, all four regimes also substantially increase the amount stolen conditional on stealing taking place (Sign test: $p < 0.01$). This causes an increase in bimodality, clearly visible in Fig. 1 which shows the distributions of the amounts taken for the three decisions and two sanction types: under sanctions, individuals tend to either steal everything or not steal at all.

5.2. Treatment comparisons: Propensity to steal

We compare the propensity to steal under each of the four treatments non-parametrically, first unconditionally, then controlling for whether or not a subject stole without sanctions. We then repeat this analysis with regressions, which also allows us to control for the size of sanctions at the individual level. As shown in table 1, the smallest proportion of subjects steal in N-MS *Severe* (49%) and the greatest in N-MS *Lenient* (63%). Behavior in the monetary sanction treatments lie between (MS *Lenient* - 57%; MS *Severe* - 56%). The only pairwise comparison that results in a statistically significant difference is between the two N-MS treatments (Z test $p = 0.015$).

As a simple way to control for differences in baseline levels of stealing, we define the dummy variable *deterred*, which takes the value 1 if a subject stole without sanctions but chose not to when threatened with punishment. We also define the counterpart *encouraged,* which takes the value 1 if a subject did not steal without sanctions but *did* steal with a sanctions regime in place. The proportions of subjects deterred and encouraged are shown by treatment in Fig. 2. N-MS *Severe* deterred 30% of individuals, significantly more than the other three treatments, for which the figures were 16–17% (Z tests: MS *Lenient* $p = 0.041$; MS *Severe* $p = 0.067$; N-MS *Lenient* $p = 0.015$). None of the other pairwise comparisons are statistically significant. There were no treatment differences between the proportions of subjects encouraged to steal, which ranged from 4–7%.

To support the non-parametric analysis and examine the impact of the size of sanctions on deterrence, we estimate probit models on the probability of stealing, clustering standard errors at the subject level (Table 2). So that we can account for differing baseline levels of stealing across individuals when comparing treatments, we only include stealing decisions under sanctions as observations, and use the decision whether to steal with no sanctions as a control variable.

The first model only includes treatment dummies (MS *Lenient* is the comparison group) and finds the only statistically significant difference to be between the two N-MS treatments ($p = 0.001$). The second model controls for whether the subject stole when not facing sanctions. Stealing without sanctions increases the probability of stealing under sanctions by around 50 percentage points ($p < 0.001$). We now find that N-MS *Severe* reduces the propensity to steal more than both MS *Lenient* ($p = 0.078$) and N-MS *Lenient* ($p < 0.01$).

In the third model, we also control for the size of the sanctions. The variable *sanctions* is the average of the largest amount where a subject preferred to place more sliders and the smallest amount where they preferred to pay in the *wtp* elicitation. This is the midpoint of the elicited interval within which the true *wtp* should lie and the size of the fine faced by subjects in the MS treatments. This coefficient is highly significant ($p < 0.01$) and implies that a subject facing the highest sanction (6.25) is around 33 percentage points less likely to steal than a subject facing the lowest sanction (0.25). In this model, N-MS *Severe* demonstrates greater deterrence than MS *Lenient* ($p = 0.053$), MS *Severe* ($p = 0.078$), and N-MS *Lenient* ($p < 0.01$). There are no other statistically significant differences between treatments.

Controlling for gender, age, risk-aversion, and cognitive ability makes little difference to the results: none of these variable are statistically significant, and their inclusion does not substantially change the magnitude of our treatment dummies (this model is thus not reported). We comment on Model 4 in Section 5.4.

We summarize our findings so far in the following results:

Result 1: Proportion of subjects stealing:

a The lowest rate of stealing occurs with a severe procedure and non-monetary sanctions.
b There is no evidence of a difference in the proportion of subjects stealing between a lenient procedure with non-monetary sanctions, a lenient procedure with monetary sanctions, or a severe procedure with monetary sanctions.

5.3. Treatment comparisons: Amount stolen

When comparing the amounts stolen in each treatment, we find a significant difference in distributions only between N-MS *Severe* (2.43) and N-MS *Lenient* (3.01) (Wilcoxon signed-rank test: $p = 0.073$).[13] This difference is not significant using a sign test ($p = 0.134$), so we do not have evidence that the difference is one of central tendency. As with the decision to steal, for comparisons between sanction types, we need to control for different baseline rates of stealing, so we compared the distributions of the differences between the amount a subject stole without sanctions and what they stole with sanctions. The distribution

[12] All tests using binary data in this paper are exact unconditional Z tests for either matched (Suissa and Shuster, 1991) or unmatched pairs (Suissa and Shuster, 1985), as appropriate. For brevity, we simply refer to them as Z tests.

[13] Details of tests for all treatment comparisons can be found in appendix A.2.

Table 3
Regression analysis of the amount stolen (OLS)

Variables	(1) *Amount*	(2) *Amount*	(3) *Amount*	(4) *Amount|Steal*	(5) *Amount|Steal*	(6) *Amount|Steal*
MS*severe*	-0.150	-0.150	-0.150	-0.151	-0.135	-0.134
	(0.261)	(0.261)	(0.264)	(0.165)	(0.163)	(0.163)
N-MS*lenient*	0.371	0.110	0.217	0.148	0.0687	0.0667
	(0.397)	(0.323)	(0.327)	(0.178)	(0.156)	(0.156)
N-MS*severe*	-0.202	-0.462	-0.355	0.387**	0.312**	0.305**
	(0.408)	(0.355)	(0.356)	(0.151)	(0.135)	(0.134)
Amount (No sanction)		0.605***	0.626***		0.183***	0.190***
		(0.0646)	(0.0635)		(0.0579)	(0.0587)
Sanction size			-0.231**			-0.0324
			(0.105)			(0.0430)
Risk aversion			0.147*			0.00828
			(0.0870)			(0.0295)
Male			0.0270			0.147
			(0.350)			(0.148)
Age			-0.0327			-0.0308*
			(0.0346)			(0.0176)
CRT			0.0543			-0.0660
			(0.150)			(0.0553)
Constant	2.636***	1.017***	1.517*	4.612***	3.948***	4.766***
	(0.288)	(0.276)	(0.895)	(0.151)	(0.308)	(0.352)
Observations	292	292	292	164	164	164
R^2	0.009	0.296	0.336	0.062	0.216	0.257
Treatment comparisons						
MS*S* vs N-MS*L*	1.78	0.64	1.26	2.73	1.64	1.58
MS*S* vs N-MS*S*	0.02	0.76	0.33	12.31***	10.56***	9.25***
N-MS*L* vs N-MS*S*	4.73**	4.71**	4.63**	6.27**	6.96***	7.56***

Notes: amount (no sanction) = amount stolen in decision without sanctions; sanction size = fine or wtp in Euros; CRT = # of correct decisions in Cognitive Reflection Test. *Treatment Comparisons* are F-statistics from tests of equality of coefficients of the treatment dummies. Standard errors clustered by subject in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4
Proportion of subjects deterred.

	All subjects	Risk-averse	Risk-loving
MS lenient	0.157	0.157	0.158
MS severe	0.171	0.216	0.053
N-MS lenient	0.171	0.105	0.237
N-MS severe	0.303	0.184	0.421

of the changes in stealing in N-MS *Severe* (average = -0.67) is significantly different than in both MS *Lenient* (average = -0.04; WMW $p = 0.070$) and MS *Severe* (average = -0.19; WMW $p = 0.098$). The results for the comparison between the N-MS treatments remain the same because it is a within-subject comparison, so the amount stolen without sanctions is the same for each treatment, leading to no change in rank ordering. No other pairwise comparison of distribution is significant, and stochastic inequality tests[14] find no differences between any treatments at conventional levels of significance.

The first two columns of Table 3 show the results of linear regressions of the amount stolen on treatment dummies, the second controlling for the amount stolen in the decision without sanctions. In both regressions, the only statistical differences between coefficients suggest that people steal less in N-MS *Severe* than in N-MS *Lenient*, which is significant in both regressions (column 1: $p = 0.031$; column 2: $p = 0.032$). Controlling for the size of sanctions, risk aversion, gender, age, and cognitive ability (column 3) does not alter this conclusion. The significance level of all the treatment comparisons remains the same. Bigger sanctions reduce the amount stolen but, as will be shown below, this effect is on the extensive rather than intensive margin. Risk aversion is marginally significant, the sign of the coefficient suggesting that more risk averse people steal more.

So far, the results are somewhat equivocal. It appears that the distribution of the amount stolen in the N-MS *Severe* treatment differs from the other treatments, but the evidence that the amount stolen is *lower* is weak. This may be because the lower rates of stealing in this treatment are counterbalanced by larger amounts being stolen. We investigate this now by comparing amounts stolen conditional on stealing having taken place.

Restricting attention to subjects who stole, Mann-Whitney tests find differences in distributions between N-MS *Severe* and each of the other three treatments (see Table 9).[15] Stochastic inequality tests show that the between-subject differences are both in the direction of subjects in the N-MS *Severe* stealing more. Looking at the differences between amounts stolen with and without sanctions finds no significant treatment differences besides a weakly significant difference between the two N-MS treatments (see Table 10). This is most likely because amounts stolen with sanctions are all close to the upper bound, so the variation between treatments is swamped by the random variation in the amount stolen without sanctions.

The linear regressions reported in the fourth and fifth columns of Table 3 find that, conditional on stealing, subjects in N-MS *Severe* steal more than in each of the other treatments, regardless of controlling for the amount stolen without sanctions. All these differences are significant at the 5% level or higher. Again, adding controls (column 6) does not alter any of these conclusions. The size of sanction is no longer significant, demonstrating that this variable has an impact only on the

[14] Without restricting the domain of distributions considered under the alternative hypothesis, one can only conclude from a rejection in the Mann-Whitney test that two distributions differ, not that one is in any sense greater than the other (Schlag, 2015). For this reason, when we find a significant difference using a Mann-Whitney test, we also perform a stochastic inequality test, which allows us to infer a directional difference (i.e. that a random draw from one treatment is likely to be higher than a random draw from another).

[15] Note that here, for the within-subject tests, we can only use data from subjects who stole in both treatments, substantially reducing our sample size.

extensive margin, as one would expect from a rational perspective. Age is marginally significant, with older subjects stealing smaller amounts.

We summarize the findings from this section in the following two results:

Result 2: Amount stolen:

a The distribution of amounts stolen in N-MS *Severe* differs weakly from the other three treatments.
b There is no evidence of a difference in the distributions of amounts stolen between a lenient procedure with non-monetary sanctions, a lenient procedure with monetary sanctions, or a severe procedure with monetary sanctions.

Result 3: Amount stolen conditional on stealing:

a Subjects who steal, steal more in N-MS *Severe* than in the other three treatments.
b There is no evidence of a difference in the distributions of amounts stolen conditional on stealing between a lenient procedure with non-monetary sanctions, a lenient procedure with monetary sanctions, or a severe procedure with monetary sanctions.

5.4. Further results

So far, we have found, contrary to standard theoretical predictions, that the different types of sanctions and procedures we consider can affect the propensity to steal and amount stolen at both the intensive and extensive margin. As suggested in Section 3, the effectiveness of theoretically equivalent sanction regimes may be affected by risk-aversion among subjects and the strength of moral condemnation signalled by the different regimes. These factors may even both be present but work in different directions, resulting in no observed difference between the two regimes. In this section, we attempt to shed some light on what might be driving our results.

Table 4 reports the proportion of subjects deterred from stealing by treatment and according to whether their answers to the incentivised risk-elicitation indicated they were risk-averse (MS: 51; N-MS: 38) or risk-loving (MS: 19; N-MS: 38). According to the theoretical predictions for MS, risk-averse subjects should be less deterred by a severe procedure than a lenient procedure, whereas the opposite should be true for risk-loving subjects. Although none of the differences are statistically significant, the data is completely contrary to the theory, with the severe procedure increasing the proportion of deterrence amongst the risk-averse (from 0.157 to 0.216) and decreasing it for the risk-loving (from 0.158 to 0.053). We view this as suggestive evidence that risk preferences are not playing an important role in driving our results.

However, the direction of treatment differences are almost entirely consistent with the hypothesized effect of changes in the strength of signals of social unacceptability between lenient versus severe procedures and monetary versus non-monetary sanctions: severe procedures deter more than lenient ones, and N-MS deters more than MS (the only comparison for which this is not true is between lenient and severe procedures for risk-loving subjects in the MS treatment).

Finally, we investigate whether the marginal effect of sanctions differs between treatments. This is tested in the final model in Table 2 where we add interaction terms between *sanctions* and treatment dummies. None of the interaction terms are statistically significant, and the joint hypothesis that they are all identical to zero cannot be rejected ($p = 0.962$). Furthermore, no interaction term is found to be significant when added individually to model 3 (not reported).

In the standard deterrence model, crime is reduced only through the increased utility cost of sanctions, and any difference in deterrence under different regimes should show up through the marginal effect of sanctions. The fact that N-MS *Severe* reduces stealing relative to the other treatments, but not through the marginal impact of sanctions, suggests to us that the difference is more likely due to this treatment signalling greater social condemnation.[16]

6. Discussion and conclusion

The theory of optimal deterrence that started with Becker's (1968) paper is the origin of major policy changes in criminal law and criminal procedures across many western countries. However, some of the most important predictions of this theory have proven to be particularly difficult to test (Levitt and Miles, 2007), and the most robust results of the related empirical literature seem to invite some skepticism about its applicability (Chalfin and McCrary, 2017). This paper provides a rigorous experimental test of the theory of optimal deterrence, focusing on the nature of the sanction and the procedure type, for which there is insufficient previous work. In the experiment, we have elicited the willingness to pay (*wtp*) to avoid a real effort task for every subject. This same task was later used in our 'hard labor' treatment as an N-MS: every convicted subject had to perform the effort task before leaving the lab. By eliciting the *wtp*, we know the monetary equivalent of the N-MS we impose on each individual and, in the monetary treatment, we impose a fine equal to their *wtp* to avoid the task.

Our first result shows that with a severe adjudicative procedure, N-MSs are more deterring than MSs. This result, however, concerns only the extensive margin: the proportion of subjects who opt to steal. Our second result examines the intensive margin (i.e. the amount stolen) and finds that neither the type of sanction nor the type of procedure seems to produce any noticeable treatment effect. However, this is because the deterring effect of N-MS *Severe* on the extensive margin (fewer subjects steal) is offset by the opposite encouraging effect on the extensive margin (those fewer subjects steal more on average), which is our third result. Furthermore, we show that risk preferences do not explain the results. Taken together, these results suggest that when replacing MSs with N-MSs of the same monetary equivalent, an authority conveys a different message to subjects, a message that induces more of them to switch their behavior from crime to law abidance. To understand why, perhaps we should look beyond deterrence. Legal and economics scholars have long recognized that the law also has an expressive function in as much as it offers guidance on the behavior that society expects individuals to follow (see Cooter 1998; Funk 2007; Nance 1997; Sunstein 1996 for a sample of the literature). Moreover, D'Antoni and Galbiati (2007) notes that the potential use of MSs as a means to increase the fiscal budget makes the guidance function of sanctions less credible: this commitment problem of the authority can be solved by implementing N-MSs.

Regarding comparisons between severe and lenient adjudicative procedures with equivalent levels of accuracy, our finding that there is no difference in the resulting level of crime when sanctions are monetary in nature is in line with the results of Markussen et al. (2016) but contradict those of Rizzolli and Stanca (2012). That a severe procedure appears to be more deterring than a lenient one in the context of N-MSs might again be tentatively explained by theories that refer to the expressive function of the law: a severe regime, by which the authority is willing to sacrifice innocents to catch all the guilty, signals that the authority views the crime to be so serious that it must be avoided at all costs. Conversely, a lenient procedure sends a much softer signal about the gravity of the crime. Clearly, further work is needed to examine the robustness of our findings and provide more conclusive evidence concerning why theoretically equivalent sanction regimes may or may not be more effective at reducing crime.

We conclude by addressing one of the recurring objections of our colleagues in Law and Economics (a field where laboratory

[16] A possible explanation for the apparent effectiveness of N-MS *Severe* would be subjects systematically under-reporting their *wtp*. However, in this case, we should also find N-MS *Lenient* more effective than the MS treatments.

experimentation is not a commonly used methodology) who are sometimes perplexed by the limited realism of lab experiments testing concepts such as ‘crime’ and ‘hard labor.’ Indeed, everything from the negligible monetary sanction amounts to the unusual nature of the non-monetary sanction and from the artificial anonymity of the environment to the stochastic detection mechanism and large error probabilities constitutes an important departure from an irreducibly complex reality. However, as experimentalists, we are more concerned with the external rather than ecological validity (Fréchette, 2015) of our conclusions: the ambition of our experiment is simply to test two hypotheses related to the effective implementation of optimal deterrence theory rather than promote any specific policy. If the deterrence value of N-MSs and MSs, or lenient and severe procedures, differ in a controlled laboratory setting, there is no reason to believe deterrence should be identical in the field.

Acknowledgement

We would like to thank Thomas Markussen, Jean-Robert Tyran, Tim Friehe, Margherita Saraceno and Matteo Migheli for useful discussions. We would also like to thank participants at the 2016 Society for Institutional and Organizational Economics annual meeting in Paris, 2015 European Law and Economics Association Annual meeting in Vienna, the 2015 Società Italiana degli Economisti annual meeting in Naples and the 2015 Italian Society of Law and Economics in Naples. The usual disclaimers apply.

Appendix A. Further results

A1. Order effects

Table 5
Order effects: amount stolen

Treatment	First choice	Second choice	WMW *p*-value
N-MS lenient	2.70	3.32	0.171
N-MS severe	2.23	2.63	0.494
M-S lenient	2.41	2.86	0.500
M-S severe	2.54	2.44	0.890

Table 6
Order effects: proportion stealing

Treatment	First choice	Second choice	z-test *p*-value
N-MS lenient	0.58	0.68	0.528
N-MS severe	0.45	0.53	0.567
M-S lenient	0. 54	0.60	0.699
M-S severe	0. 57	0.54	0.897

Table 7
Amount stolen - Lower left: comparisons of distributions (within-subject: Wilcoxon Signed Rank-sum; between subject: WMW); Upper right: comparisons of stochastic inequality (within-subject: Sign test; between subject: Stochastic inequality test).

	M-S lenient	M-S severe	N-MS lenient	N-MS severe
M-S lenient		n.s.	n.s.	n.s.
M-S severe	-0.15		n.s.	n.s.
N-MS lenient	0.37	0.52		n.s.
N-MS severe	-0.20	-0.05	-0.57*	

Table 8
Change in the amount stolen from decision without sanctions - Lower left: comparisons of distributions (within-subject: Wilcoxon Signed Rank-sum; between subject: WMW); Upper right: comparisons of stochastic inequality (within-subject: Sign test; between subject: Stochastic inequality test).

	M-S lenient	M-S severe	N-MS lenient	N-MS severe
M-S lenient		n.s.	n.s.	n.s.
M-S severe	-0.15		n.s.	n.s.
N-MS lenient	-0.06	0.09		n.s.
N-MS severe	-0.63*	-0.48*	-0.57*	

Table 9
Amount stolen conditional on stealing - Lower left: comparisons of distributions (within-subject: Wilcoxon Signed Rank-sum; between subject: WMW); Upper right: comparisons of stochastic inequality (within-subject: Sign test; between subject: Stochastic inequality test). Within subject tests: n = 32 (M-S); n = 33 (NM-S).

	M-S lenient	M-S severe	N-MS lenient	N-MS severe
M-S lenient (n = 39)		n.s.	n.s.	**
M-S severe (n = 40)	-0.07		n.s.	***
N-MS lenient (n = 48)	0.15	0.1035		n.s.
N-MS severe (n = 37)	0.39***	0.54***	0.11*	

Table 10
Lower left: comparisons of distributions (within-subject: Wilcoxon Signed Rank-sum; between subject: WMW); Upper right: comparisons of stochastic inequality (within-subject: Sign test; between subject: Stochastic inequality test). Within subject tests: n = 32 (M-S); n = 33 (NM-S).

	M-S lenient	M-S severe	N-MS lenient	N-MS severe
M-S lenient (n = 39)		n.s.	n.s.	n.s.
M-S severe (n = 40)	-0.07		n.s.	n.s.
N-MS lenient (n = 48)	-0.28	-0.22		n.s.
N-MS severe (n = 37)	-0.02	0.04	0.11*	

A2. Treatment effects

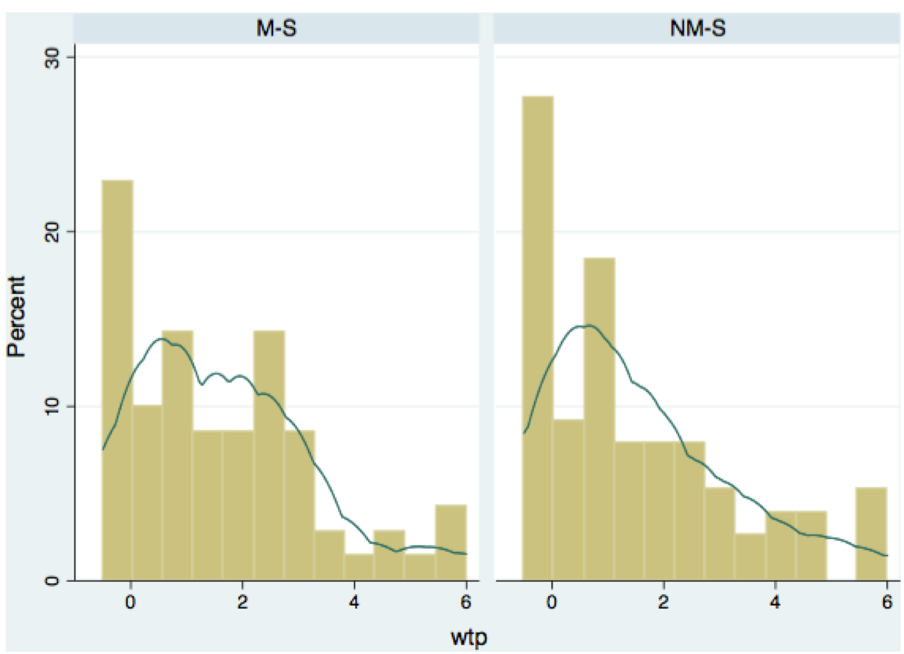


Fig. 3. Distribution of *wtp* by treatment

Appendix B. Instructions & screenshots

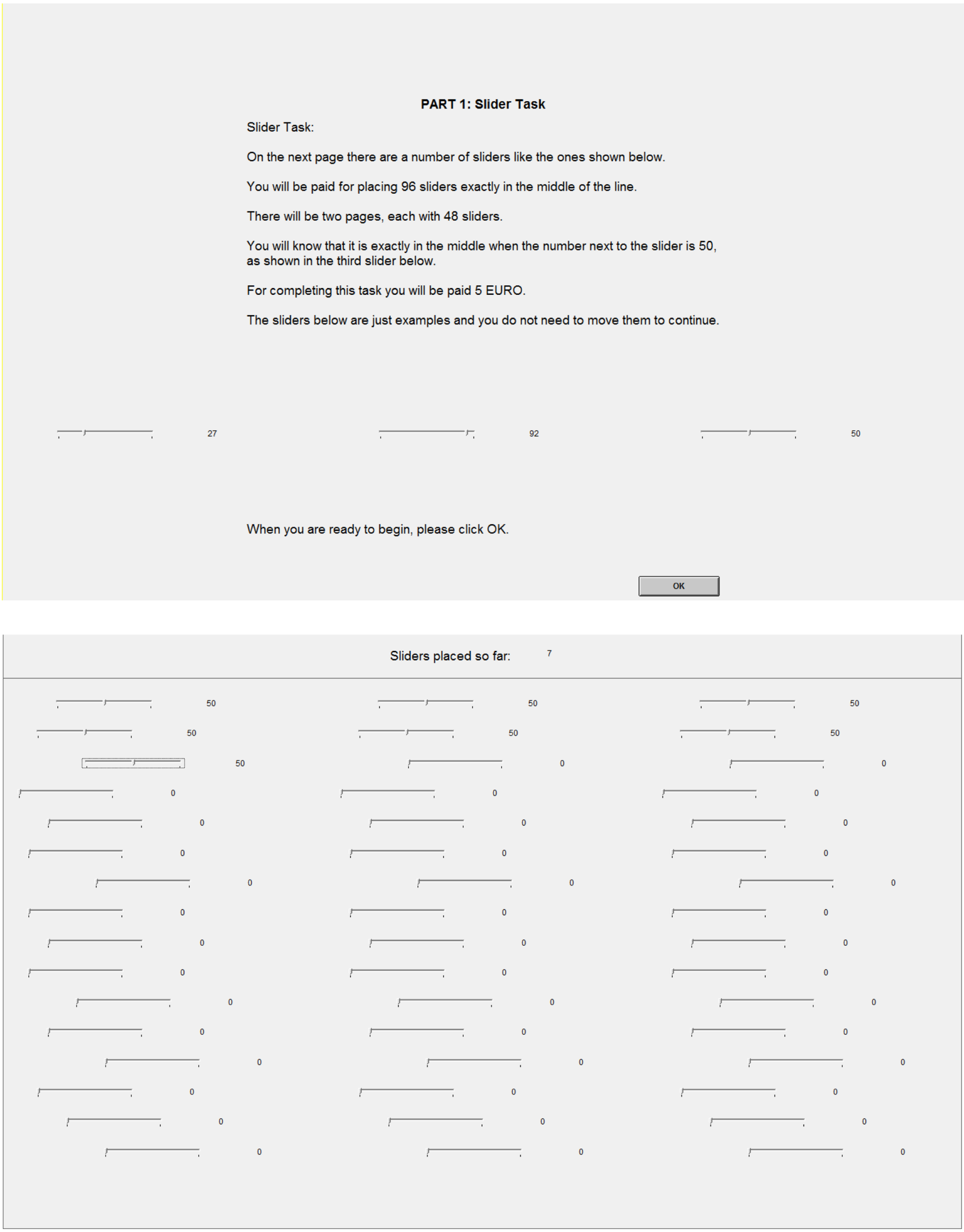


Fig. 4. Instructions for the slider task

PART 1: Slider Task

You will now be given an extra 6 EURO. You may use some or all of this money in the following task (the money you do not spend will be added to your final earnings) :

The table below gives a list of 13 choices between correctly placing a further 96 sliders and different amounts of money you can pay to avoid having to do this.

At the end of the experiment, one of these 13 choices will be randomly selected.

If, for the randomly selected choice, you chose to do the sliders, then you must again correctly place 96 sliders before collecting the money you have earned in this experiment.

If, for the randomly selected choice, you chose the payment, you will be able to collect your money immediately, but the payment will be deducted from the amount you receive.

You will not receive extra money for any further sliders you place, but only the money you have already earned.

To indicate your responses, please place a checkmark in the corresponding box.

Note that your responses must be consistent, so if you say you prefer to pay a certain amount to avoid placing further sliders, you must also say you prefer to pay all lesser amounts.

Would you prefer to place a further 96 sliders or pay 0 EURO?	☐ Repeat the slider task	☐ Pay 0 EURO
Would you prefer to place a further 96 sliders or pay 0.50 EURO?	☐ Repeat the slider task	☐ Pay 0.5 EURO
Would you prefer to place a further 96 sliders or pay 1.00 EURO?	☐ Repeat the slider task	☐ Pay 1 EURO
Would you prefer to place a further 96 sliders or pay 1.50 EURO?	☐ Repeat the slider task	☐ Pay 1.5 EURO
Would you prefer to place a further 96 sliders or pay 2.00 EURO?	☐ Repeat the slider task	☐ Pay 2 EURO
Would you prefer to place a further 96 sliders or pay 2.50 EURO?	☐ Repeat the slider task	☐ Pay 2.5 EURO
Would you prefer to place a further 96 sliders or pay 3.00 EURO?	☐ Repeat the slider task	☐ Pay 3 EURO
Would you prefer to place a further 96 sliders or pay 3.50 EURO?	☐ Repeat the slider task	☐ Pay 3.5 EURO
Would you prefer to place a further 96 sliders or pay 4.00 EURO?	☐ Repeat the slider task	☐ Pay 4 EURO
Would you prefer to place a further 96 sliders or pay 4.50 EURO?	☐ Repeat the slider task	☐ Pay 4.5 EURO
Would you prefer to place a further 96 sliders or pay 5.00 EURO?	☐ Repeat the slider task	☐ Pay 5 EURO
Would you prefer to place a further 96 sliders or pay 5.50 EURO?	☐ Repeat the slider task	☐ Pay 5.5 EURO
Would you prefer to place a further 96 sliders or pay 6.00 EURO?	☐ Repeat the slider task	☐ Pay 6 EURO

FINISH

Fig. 5. Instructions for the *wtp* elicitation task

PART 2: Choices

In this part of the experiment you will make 3 choices.

For each choice, you will be asked how much money you want to take from another participant.

The possible consequences of your decision will be different for each of the 3 choices, as will be explained in detail before you make each choice.

Determination of your payment for Part 2

Whether or not one of your choices is implemented is determined randomly with equal probability. **EITHER**:
- **one** of your three choices will be randomly chosen to be implemented, and you will receive the amount you have chosen to take, and another participant in the experiment will have the same amount deducted from their final payment, **OR**
- **one** of another participant's three choices will be randomly chosen to be implemented, and you will have the amount they chose to take from you deducted from your final payment.

All choices are strictly anonymous.

Please click Continue when you are ready to make your first choice.

Continue

Fig. 6. General instructions for stealing decisions

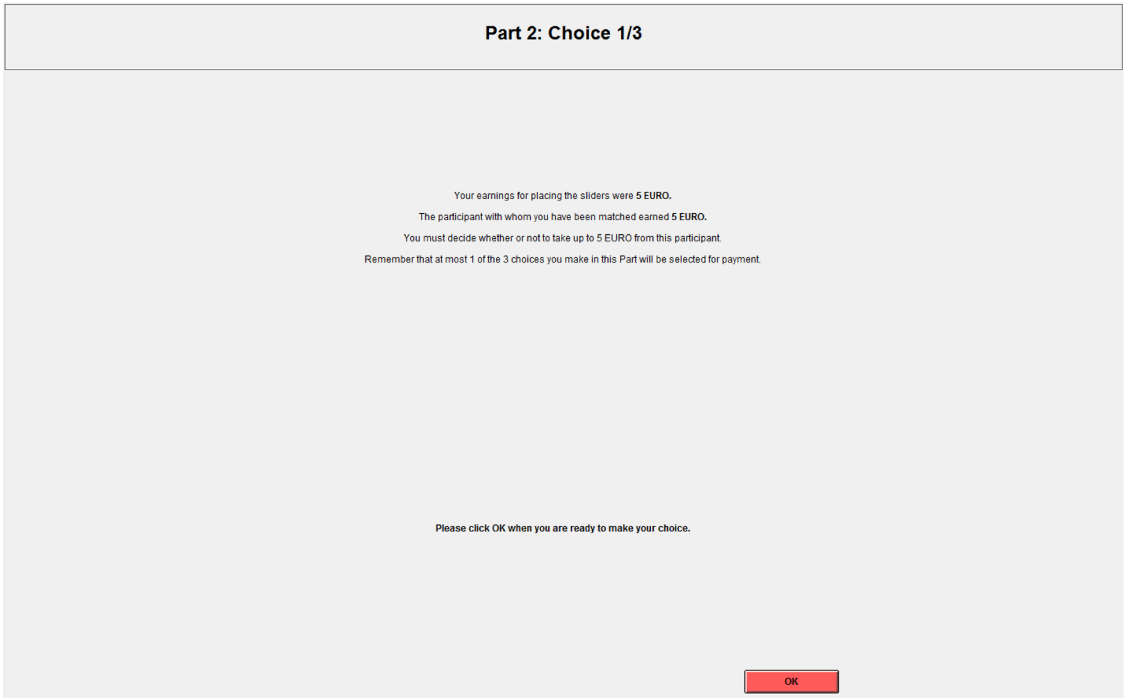


Fig. 7. Instructions for the stealing decisions in the *No-deterrence* procedure. These instructions were the same for both MS and N-MS treatments

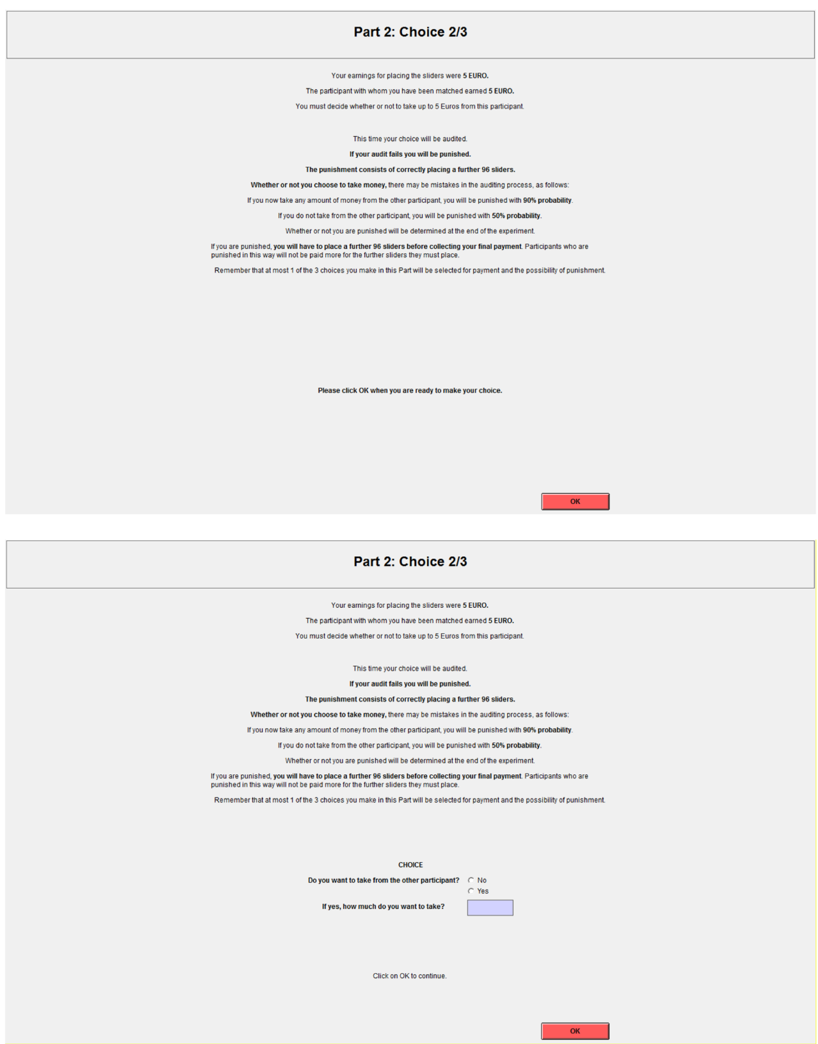
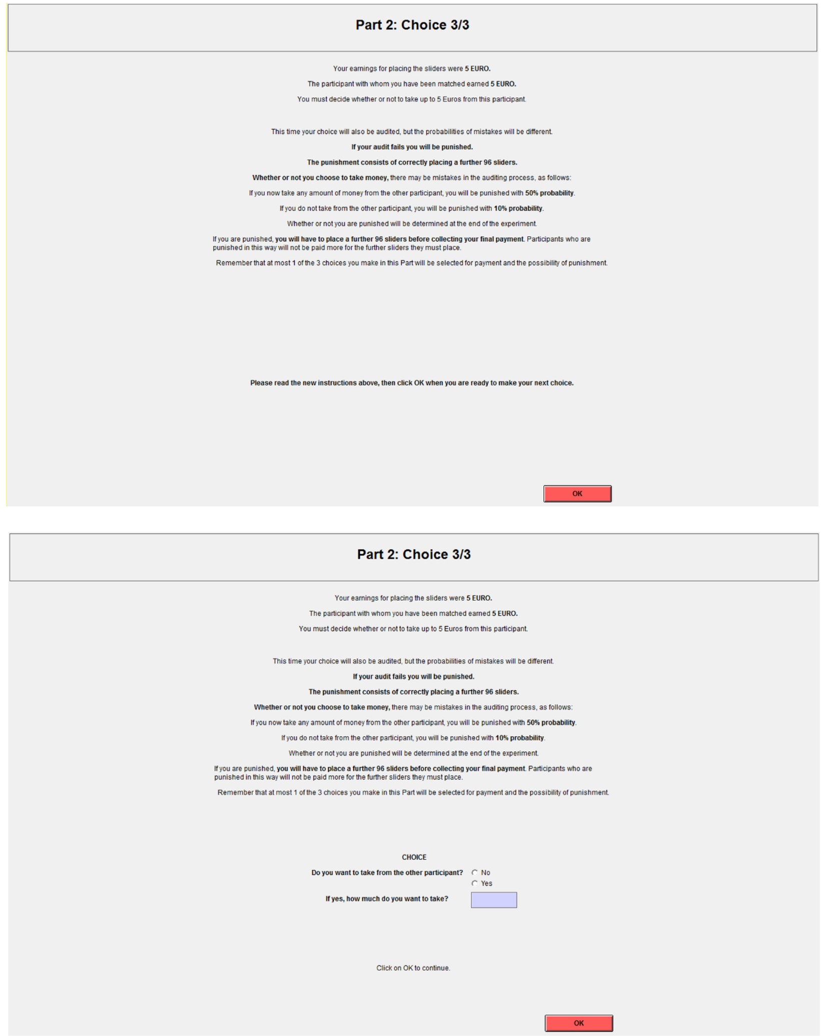


Fig. 8. Instructions for the stealing decisions with a *Severe* procedure. The instructions for the MS treatment are above, and the instructions for the N-MS treatment are below.



Part 2: Choice 3/3

Your earnings for placing the sliders were **5 EURO.**

The participant with whom you have been matched earned **5 EURO.**

You must decide whether or not to take up to 5 Euros from this participant.

This time your choice will also be audited, but the probabilities of mistakes will be different.

If your audit fails you will be punished.

The punishment consists of correctly placing a further 96 sliders.

Whether or not you choose to take money, there may be mistakes in the auditing process, as follows:

If you now take any amount of money from the other participant, you will be punished with **50% probability**.

If you do not take from the other participant, you will be punished with **10% probability**.

Whether or not you are punished will be determined at the end of the experiment.

If you are punished, **you will have to place a further 96 sliders before collecting your final payment**. Participants who are punished in this way will not be paid more for the further sliders they must place.

Remember that at most 1 of the 3 choices you make in this Part will be selected for payment and the possibility of punishment.

Please read the new instructions above, then click OK when you are ready to make your next choice.

OK

Part 2: Choice 3/3

Your earnings for placing the sliders were **5 EURO.**

The participant with whom you have been matched earned **5 EURO.**

You must decide whether or not to take up to 5 Euros from this participant.

This time your choice will also be audited, but the probabilities of mistakes will be different.

If your audit fails you will be punished.

The punishment consists of correctly placing a further 96 sliders.

Whether or not you choose to take money, there may be mistakes in the auditing process, as follows:

If you now take any amount of money from the other participant, you will be punished with **50% probability**.

If you do not take from the other participant, you will be punished with **10% probability**.

Whether or not you are punished will be determined at the end of the experiment.

If you are punished, **you will have to place a further 96 sliders before collecting your final payment**. Participants who are punished in this way will not be paid more for the further sliders they must place.

Remember that at most 1 of the 3 choices you make in this Part will be selected for payment and the possibility of punishment.

CHOICE

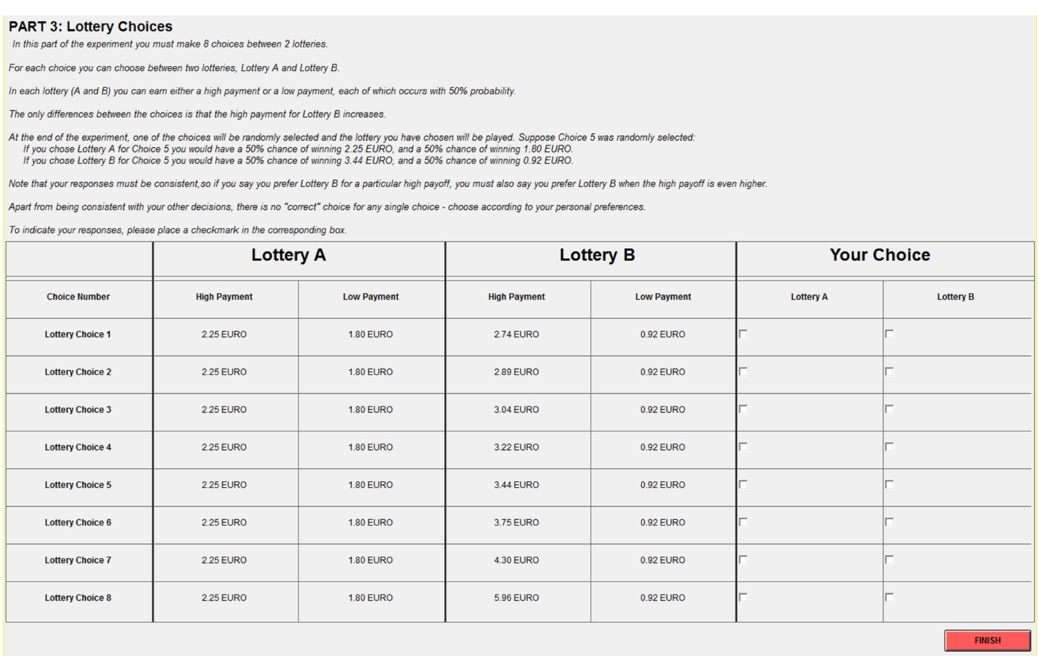
Do you want to take from the other participant? No / Yes

If yes, how much do you want to take?

Click on OK to continue.

OK

Fig. 9. Instructions for the stealing decision in the *Lenient* procedure. The instructions for the MS treatment are above, and the instructions for the N-MS treatment are below.



PART 3: Lottery Choices

In this part of the experiment you must make 8 choices between 2 lotteries.

For each choice you can choose between two lotteries, Lottery A and Lottery B.

In each lottery (A and B) you can earn either a high payment or a low payment, each of which occurs with 50% probability.

The only differences between the choices is that the high payment for Lottery B increases.

At the end of the experiment, one of the choices will be randomly selected and the lottery you have chosen will be played. Suppose Choice 5 was randomly selected:
If you chose Lottery A for Choice 5 you would have a 50% chance of winning 2.25 EURO, and a 50% chance of winning 1.80 EURO.
If you chose Lottery B for Choice 5 you would have a 50% chance of winning 3.44 EURO, and a 50% chance of winning 0.92 EURO.

Note that your responses must be consistent,so if you say you prefer Lottery B for a particular high payoff, you must also say you prefer Lottery B when the high payoff is even higher.

Apart from being consistent with your other decisions, there is no "correct" choice for any single choice - choose according to your personal preferences.

To indicate your responses, please place a checkmark in the corresponding box.

	Lottery A		Lottery B		Your Choice	
Choice Number	High Payment	Low Payment	High Payment	Low Payment	Lottery A	Lottery B
Lottery Choice 1	2.25 EURO	1.80 EURO	2.74 EURO	0.92 EURO		
Lottery Choice 2	2.25 EURO	1.80 EURO	2.89 EURO	0.92 EURO		
Lottery Choice 3	2.25 EURO	1.80 EURO	3.04 EURO	0.92 EURO		
Lottery Choice 4	2.25 EURO	1.80 EURO	3.22 EURO	0.92 EURO		
Lottery Choice 5	2.25 EURO	1.80 EURO	3.44 EURO	0.92 EURO		
Lottery Choice 6	2.25 EURO	1.80 EURO	3.75 EURO	0.92 EURO		
Lottery Choice 7	2.25 EURO	1.80 EURO	4.30 EURO	0.92 EURO		
Lottery Choice 8	2.25 EURO	1.80 EURO	5.96 EURO	0.92 EURO		

FINISH

Fig. 10. Instructions for the risk elicitation mechanism

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.socec.2018.09.011

References

Avio, K.L., 1993. Economic, retributive and contractarian conceptions of punishment. Law and Philosophy 12 (3), 249–286.
Becker, G., 1968. Crime and punishment: An economic approach. J. Politic. Econ. 76, 169–217.
Beckett, K., Harris, A., 2011. On cash and conviction. Criminol. Public Policy 10 (3), 509–537.
Blais, A.-R., Weber, E.U., 2006. A domain-specific risk-taking (dospert) scale for adult populations. Judg. Decision Making 1 (1).
Bowles, S., Polania-Reyes, S., 2012. Economic incentives and social preferences: substitutes or complements? J. Econ. Lit. 50 (2), 368–425.
Chalfin, A., McCrary, J., 2017. Criminal deterrence: A review of the literature. J. Econ. Lit. 55 (1), 5–48.
Cinyabuguma, M., Page, T., Putterman, L., 2005. Cooperation under the threat of expulsion in a public goods experiment. J. Pub. Econ. 89 (8), 1421–1435.
Cooter, R., 1998. Expressive law and economics. J. Legal Stud. 27 (2), 585–608.
Csermely, T., Rabas, A., 2016. How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods. J.RiskUncertain. 53 (2-3), 107–136.
D'Antoni, M., Galbiati, R., 2007. A signalling theory of non-monetary sanctions. Intern. Rev. Law Econ. 27 (2), 204–218.
Dhami, S., al Nowaihi, A., 2013. An extension of the becker proposition to non-expected utility theory. Math. Soc. Sci. 65 (1), 10–20.
Drichoutis, A., Lusk, J., 2012. What can multiple price lists really tell us about risk preferences?.
Edmonds, D., Duncan, S., Armstrong, G., 2014. Would You Kill the Fat Man? Recorded Books, Incorporated.
Ehrlich, I., 1981. On the usefulness of controlling individuals: an economic analysis of rehabilitation, incapacitation and deterrence. Am. Econ. Rev. 307–322.
Fiedler, M., Haruvy, E., 2017. The effect of third party intervention in the trust game. J. Behav. Exp. Econ. 67, 65–74.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp.Econ. 10 (2), 171–178.
Fréchette, G. R., 2015. Handbook of Experimental Economic Methodology. Oxford University Press. chapter Laboratory experiments: Professionals versus students.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ.Perspect. 25–42.
Friedman, D.D., 1981. Reflections on optimal punishment, or: should the rich pay higher fines? Res. Law Econ. 3, 185–205.
Funk, P., 2007. Is there an expressive function of law? an empirical analysis of voting laws with symbolic fines. Am. Law Econ. Rev.
Garoupa, N., 1997. The theory of optimal law enforcement. J. Econ. Surv. 11 (3), 267–295.
Garoupa, N., 2003. Behavioral economic analysis of crime: A critical review. Eur. J. Law Econ. 15 (1), 5–15.
Garoupa, N., 2014. Economic theory of criminal behavior. Encyclopedia of Criminology and Criminal Justice. Springer, pp. 1280–1286.
Garoupa, N., Klerman, D., 2004. Corruption and the optimal use of nonmonetary sanctions. Intern. Rev. Law Econ. 24 (2), 219–225.
Garoupa, N., Rizzolli, M., 2013. Wrongful convictions do lower deterrence. J. Inst. Theor. Econ. 168, 224–231.
Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. Am. Econ. Rev. 102 (1), 469–503.
Gneezy, U., Rustichini, A., 2000. Fine is a price, a. J. Legal Stud. 29, 1.
Grant, R.W., 2018. Incentives and praise compared: the ethics of motivation. Intern. Rev. Econ. 1–12.
Grechenig, K., Nicklisch, A., Thöni, C., 2010. Punishment despite reasonable doubt—a public goods experiment with sanctions under uncertainty. J. Emp. Legal Stud. 7 (4), 847–867.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. J. Econ. Sci. Assoc. 1 (1), 114–125.
Harris, A., Evans, H., Beckett, K., 2010. Drawing blood from stones: Legal debt and social inequality in the contemporary united states1. Am. J. Soc. 115 (6), 1753–1799.
Kahan, D.M., 1996. What do alternative sanctions mean? Univ. Chicago Law Rev. 63 (2), 591–653.
Kahan, D.M., 2005-2006. What's really wrong with shaming sanctions. Texas Law Rev. 84, 2075–2096.
Kahan, D.M., Posner, E.A., 1999. Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines. J. Law Econ. 42 (1), 365–391.
Kan, S.S., 1996. Corporal punishments and optimal incapacitation. J. Legal Stud. 25 (1), 121–130.
Lando, H., 2006. Does wrongful conviction lower deterrence? J. Legal Stud. 35 (2), 327–338.
Lando, H., Mungan, M.C., 2017. The effect of type-1 error on deterrence. Intern. Rev. Law Econ.
Levitt, S.D., 1997. Incentive compatibility constraints as an explanation for the use of prison sentences instead of fines. Intern. Rev. Law Econ. 17 (2), 179–192.
Levitt, S.D., 1997. Using electoral cycles in police hiring to estimate the effect of police on crime. Am. Econ. Rev. 87 (3), 270–290.
Levitt, S.D., Miles, T.J., 2007. Empirical study of criminal punishment. Handbook LawEcon. 1, 455–495.
Marchegiani, L., Reggiani, T., Rizzolli, M., 2016. Loss averse agents and lenient supervisors in performance appraisal. J. Econ. Behav. Org. 131, 183–197.
Markel, D., Flanders, C., 2010. Bentham on stilts: The bare relevance of subjectivity to retributive justice. California Law Rev. 98, 907–988.
Markussen, T., Putterman, L., Tyran, J.-R., 2016. Judicial error and cooperation. Eur. Econ. Rev. 89, 372–388.
Masclet, D., Noussair, C., Tucker, S., Villeval, M., 2003. Monetary and nonmonetary punishment in the voluntary contributions mechanism. Am. Econ. Rev. 366–380.
McAdams, R.H., 2000. An attitudinal theory of expressive law. Oregon Law Rev. 79, 339–390.
Miceli, T.J., 2010. A model of criminal sanctions that incorporate both deterrence and incapacitation. Econ. Lett. 107 (2), 205–207.
Montag, J., Tremewan, J., 2018. Let the punishment fit the criminal: an experimental study. J. Econ. Behav. Org.
Mungan, M., 2011. A utilitarian justification for heightened standards of proof in criminal trials. J. Inst. Theor. Econ. 167 (2).
Mungan, M.C., 2012. The law and economics of fluctuating criminal tendencies and incapacitation. Maryland Law Rev.
Murphy, K.R., Balzer, W.K., 1989. Rater errors and rating accuracy. J. Appl. Psychol. 74 (4), 619–624.
Nance, D.A., 1997. Guidance rules and enforcement rules: A better view of the cathedral. Virginia Law Rev. 83 (5), 837–937.
Nicita, A., Rizzolli, M., 2014. In dubio pro reo. behavioral explanations of pro-defendant bias in procedures. CESifo Econ. Stud. ift016.
Noussair, C., Tucker, S., 2005. Combining monetary and social sanctions to promote cooperation. Econ. Inq. 43 (3), 649–660.
Ognedal, T., 2005. Should the standard of proof be lowered to reduce crime? Intern. Rev. Law Econ. 25 (1), 45–61.
Peeters, R., Vorsatz, M., 2013. Immaterial rewards and sanctions in a voluntary contribution experiment. Econ. Inq. 51 (2), 1442–1456.
Pérez, R.L., Kiss, H.J., 2012. Do people accurately anticipate sanctions? South. Econ. J. 79 (2), 300–321.
Png, I.P.L., 1986. Optimal subsidies and damages in the presence of judicial error. Intern. Rev. Law Econ. 6 (1), 101–105.
Polinsky, A.M., Shavell, S., 1984. The optimal use of fines and imprisonment. J. Pub. Econ. 24 (1), 89–99.
Polinsky, A.M., Shavell, S., 1984. The optimal use of fines and imprisonment. J. Pub. Econ. 24 (1), 89–99.
Polinsky, A.M., Shavell, S., 2009. Criminal Law and Economics - Encyclopedia of Law and Economics. 3. Edward Elgar Pub., Amsterdam, pp. 1–60.
Posner, R.A., 1980. Retribution and related concepts of punishment. J. Legal Stud. 71–92.
Prendergast, C., 1999. The provision of incentives in firms. J. Econ. Lit. 37 (1), pp.7–63.
Rizzolli, M., 2018. Type-I and Type-II errors. In: Marciano, A., Ramello, G.B. (Eds.), Encyclopedia of Law and Economics. Springer New York, New York, NY, pp. 1–9. https://doi.org/10.1007/978-1-4614-7883-6_66-1.
Rizzolli, M., Saraceno, M., 2013. Better that ten guilty persons escape: punishment costs explain the standard of evidence. Public Choice 155 (3-4), 395–411.
Rizzolli, M., Stanca, L., 2012. Judicial errors and crime deterrence: theory and experimental evidence. J. Law Econ. 55 (2), 311–338.
Ruback, R.B., 2015. The benefits and costs of economic sanctions: Considering the victim, the offender, and society. Minnesota Law Rev. 99, 1779.
Schlag, K. H., 2015. Who gives direction to statistical testing? best practice meets mathematically correct tests.
Shavell, S., 1986. The judgment proof problem. Intern. Rev. Law Econ. 6 (1), 45–58.
Shavell, S., 2015. A simple model of optimal deterrence and incapacitation. Intern. Rev. Law Econ. 42, 13–19.
Sugden, R., 2018. Awards, incentives and mutual benefit. Intern. Rev. Econ. https://doi.org/10.1007/s12232-017-0288-1.
Suissa, S., Shuster, J.J., 1985. Exact unconditional sample sizes for the 2 × 2 binomial trial. J. Royal Stat. Soc.. Series A (General) 317–327.
Suissa, S., Shuster, J.J., 1991. The 2 x 2 matched-pairs trial: Exact unconditional design and analysis. Biomet. 361–372.
Sunstein, C.R., 1996. On the expressive function of law. Univ. Pennsylvania Law Rev. 144 (5), pp.2021–2053.
Travis, J., Western, B., Redburn, S., et al., 2014. The Growth of Incarceration in the United States:: Exploring Causes and Consequences. Nation. Acad. Press.
Volokh, A., 1997. n guilty men. Univ. Pennsylvania Law Rev. 146 (1), 173–216.
Wittman, D., 1974. Punishment as retribution. Theor. Decision 4 (3-4), 209–237.