WORKING PAPERS

Hard Labour in the lab: Are monetary and non-monetary sanctions really substitutable?

Matteo Rizzolli
James Tremewan

June 2016

Working Paper No: 1606

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Hard Labour in the lab: Are monetary and non-monetary sanctions really substitutable?*

Matteo Rizzolli[†] and James Tremewan[‡]

June 30, 2016

Abstract

The theory of optimal deterrence suggests the substitution of monetary sanctions over non-monetary sanctions whenever this is possible because non-monetary sanctions are more socially costly. This prescription is based on the assumption that monetary and non-monetary sanctions are perfect substitutes: there exists a monetary equivalent of a non-monetary sanction that, if used as a fine, produces the same level of deterrence. We test this assumption with an experiment. In our stealing game potential thieves face the possibility of punishment. Our non-monetary sanction treatments mimic hard labour: we require convicted individuals to carry out a tedious real effort task. In the monetary treatments sanctions are instead fines, which are based on individuals' willingness to pay to avoid the effort task to ensure comparability to the non-monetary treatment. A second manipulation of our experiment concerns the balance of errors in the adjudication procedure (convictions 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 is consistent with the notion that both monetary punishment and pro-defendant sanction regimes are less effective in communicating moral condemnation of an act.

Keywords: monetary sanctions, non-monetary sanctions, hard labor, optimal deterrence, experimental economics, effort task, punishment, adjudicative procedures, type-I errors, type-II errors, Leniency error, Severity error

JEL Codes: D01, K14, K40, K41, K42

*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 & Economics Association Annual meeting in Vienna, the 2015 Società Italiana degli Economisti annual meeting in Naples and the 2015 Italian Society of Law & Economics in Naples. The usual disclaimers apply.

[†]LUMSA University, Rome email: m.rizzolli@lumsa.it

[‡]University of Vienna, Vienna Center for Experimental Economics email: james.tremewan@univie.ac.at

1 Introduction

Becker's (1968) seminal work on the theory of optimal deterrence, among other things, prescribes policymakers to substitute whenever possible non-monetary sanctions (N-MSs) with monetary sanctions (MSs). The intuition is simple: if we substitute a N-MS with a MS while keeping deterrence constant, we reduce social costs because MSs are costless transfers from convicted individuals to the rest of society while N-MSs are costly for the convicted as well as for society.[1]

The general revival of the deterrence approach in the years following Becker's article had profound consequences on public enforcement policy especially in the US.[2] Since the 70s, major reforms meant at increasing deterrence have involved the reintroduction in many US states of the death penalty, mandatory minimum sentences, "tough on crime" and "three-strike" laws. As a result the number of incarcerated has grown from fewer than 200,000 prisoners in 1972 to more than 1.6 million in 2010 and, even accounting for the general population growth this implies a four fold increase in the rate of incarceration (Travis et al., 2014). Other forms of N-MSs have also been increasingly used: intermediate or alternative sanctions such as home detention (under probation or parole), mandated community service, and mandatory drug treatment. Also the use of shaming sanctions have been explored.[3] However, regarding the substitution of non-monetary sanctions with monetary ones the picture is mixed: on one hand the number and magnitude of MSs has grown dramatically, on the other hand the increased use of N-MSs has also been remarkable (Harris et al., 2010; Ruback, 2015). In other western countries where there was no such a general surge in deterrence strategies, the shift in penalties from N-MSs to MSs appears to be cleaner

[1]Of course MSs also imply social costs (the costs of collecting fines or confiscating assets if individuals resist to pay the fine) but in the literature these costs are usually ignored or normalized to zero; what matters is that they are less costly than N-MSs. On this point see Polinsky and Shavell (1992).

[2]Of course deterrence theory is much older then Becker and must go back at least to Cesare Beccaria (1763) and Jeremy Bentham (1789). With some isolated exceptions the theory remained largely minoritarian until a major revival in the late sixties of the last century when both Becker (1968) and Gibbs (1968) resurrected the field. See Paternoster (2010) for a recent interdisciplinary review of the literature on deterrence.

[3]Modern shaming penalties include publishing names of convicted on billboards, newspapers or even broadcasting them on TV; requiring the convicted to wear certain signs or at least to attach it to some of their properties (such as special plates or bumper stickers to be carried on vehicles) or imposing self-debasement through some ritual ceremony such as a public apology. Shaming sanctions have received some renewed attention as they substitute costly centralized public enforcement with third party private punishment in the form of stigma (Kahan, 1996, 2006; Kahan and Posner, 1999).

(Beckett and Harris, 2011).

Whether MSs and N-MSs are perfect substitutes is thus an important empirical question that has very relevant policy implications. With this project we aim to test experimentally the assumption that there exists a monetary equivalent of a N-MS that keeps deterrence constant by comparing subjects' propensity to steal in a laboratory setting under different sanction regimes. The experiment in a nutshell takes the following form: we first elicit subjects' willingness to pay (*wtp*) to avoid a tedious real-effort task and then, after the stealing decision is made and the adjudication process completed, impose on convicted individuals either a MS equal to their *wtp* or a N-MS in the form of an obligation to carry out the same effort task (our hard labour). In this way we can compare N-MSs with MSs that have the same monetary equivalent disutility of effort, a comparison that would be impossible outside of the laboratory. We also exogenously manipulate the conviction probabilities for both innocent and guilty individuals in order to see whether the nature of the adjudication procedure has an asymmetric impact in procedures that are more or less pro-defendant.

Our results suggest that N-MSs and MSs are not perfect substitutes; in particular, a N-MS coupled with a less pro-defendant error structure, reduces crime more than a N-MS with a more pro-defendant error structure or either of the MS procedures. All sanctions 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 less pro-defendant error structure, leading to very little difference between treatments in total amounts stolen. The data suggests 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 sanction 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 deterrence theory. The paper proceeds as follows: in section 2 we review the literature; in section 3 we present our research hypothesis and our experimental design; in section 4 we present the main results and in section 5 we conclude.

2 Literature Review

The law & economics literature starts with Becker's (1968) informal claim made on purely utilitarian ground that MSs should substitute N-MSs as much as possible. The argument is then stated formally by Friedman (1981) and Polinsky and Shavell (1984a): since MSs are costless transfers while N-MSs are not, the substitution of the latter with the former increases social welfare.

Specific exceptions exist to this general rule: N-MSs are to be preferred over MSs when some defendants are both too poor to pay the fine and thus cannot be deterred (Polinsky and Shavell, 1984b; Shavell, 1986) and too rich to be deterred at all;[4] when the defendant can hide his wealth or the wealth depends on his 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). All these exceptions justify the use of N-MSs within the deterrence framework. On the other hand retributivist theories of punishment also justify the use of N-MSs on the ground that punishment should fit the crime, not just in size (the principle of proportionality) but also in kind (Wittman, 1974; Posner, 1980; Avio, 1993). Furthermore incapacitation theories of crime justify N-MSs such as prison and banishment as they are effective means to avoid further damages to society (Ehrlich, 1981; Kan, 1996; Miceli, 2010; Mungan, 2012; Shavell, 2015).

Another route through which sanctions may reduce crime besides the reduction in utility itself, is the possibility that "law changes behavior by signaling the underlying attitudes of a community or society" (McAdams, 2000). This is also called the expressive (Cooter, 1998) or guidance (Nance, 1997) function of the law. From this point of view, 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); Bowles and Polania-Reyes (2012).

The experimental literature on punishment is vast but only few papers relate closely to ours. Masclet et al. (2003) show that an individual expression of disapproval can partly substitute monetary sanctions in a voluntary contribution mechanism (VCM) type of game. Their result has been supported by a number of papers (Noussair and Tucker, 2005; Pérez and Kiss, 2012; Peeters and Vorsatz, 2013); however, what they call N-MS is more accurately a form of second-party shaming punishment. In our project MSs and N-MSs

[4]Fines are often considered *unjust* because the rich are able to pay fines, whereas the poor serve jail sentences (Levitt, 1997a). Furthermore, for individuals with decreasing marginal utility of income the same fine implies a smaller disutility for the rich than for the poor and this may imply a further level of unfairness. This would call for making the fine proportional on income. An opposite argument can also be made that rich may suffer more than the poor from the same non-monetary sanctions (Montag and Sobek, 2014).

are both imposed by a central authority and they are also anonymous so that shaming cannot play any role. Cinyabuguma et al. (2005) also introduce a N-MS in a VCM game in the form of banishment. Several experimental papers test the deterrence hypothesis in the lab using the same inverse dictator game we use as a baseline (Harbaugh et al., 2011; Rizzolli and Stanca, 2012; Schildberg-Hörisch and Strassmair, 2012; Feess et al., 2014; Khadjavi, 2015). However all these papers deal with MSs only. The only study we are aware of that compares MSs and N-MSs is Montag and Tremewan (2016), which investigates whether people are willing to condition the level of punishment on the subjective experience of the convicted.

This paper also looks at the asymmetric effects of sanction types on type-I (wrongful convictions of innocents) and type-II errors (wrongful acquittals of guilty individuals). The standard model of optimal deterrence predicts that both type of errors are symmetrically detrimental to deterrence (see Rizzolli 2016). On the other hand several papers have put the same theoretical prediction to the experimental test. Grechenig et al. (2010) first show with a lab experiment that judicial errors greatly undermine deterrence in a voluntary contribution mechanism (VCM) type of game. Rizzolli and Stanca (2012) disentangle the effects and find that type-I errors are more detrimental to deterrence than type-II errors but they do not reject the hypothesis that risk-aversion alone could explain this asymmetry. Marchegiani et al. (2013) find the same effect in a principal-agent setting. In contrast, Markussen et al. (2014) find a symmetric effect of type-I and -II errors in a VCM framework.

With regard to the view of punishment as conveying social condemnation, a type-II error weakens the message by signalling that the state is not overly concerned about the offender's misconduct (Markel and Flanders, 2010), whereas erring on the side of type-I errors signals that an action is sufficiently unacceptable that innocents should be sacrificed in order to ensure punishment of norm-violators. This would suggest an advantage of reducing type-II errors in favour of type-I. An interesting finding from Markussen et al. (2014) is that when subjects vote for which type of error they will face, type-I errors become even less detrimental while type-II errors become more so: while 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

In the following section we present the standard model of optimal deterrence within which we manipulate two key variables for our experiment: the type of sanction and the error balance. For clarity, we refer sanction *types* (MS vs N-MS), sanction *procedure* (error balance), and sanction *regime* (the type/procedure combination). When discussing experimental results, the terms regime and treatment will be used interchangeably.

Let b be the gains from crime, 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 separable from the utility of the monetary gain from crime while this is not the case for the MS.[5] 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 type-I error (the probability that the individual abstains from committing the crime and is wrongfully sanctioned) and ε_2 be the probability of a type-II error (the probability that the individual commits the crime and is nevertheless not sanctioned). $A = 1 - \varepsilon_1 - \varepsilon_2$ is a measure of the accuracy of the adjudicative process. Notice that two very different states of the world, with either a *lenient* or a *severe* procedure,[6] can be characterized by the same level of accuracy if $A^{LENIENT} = \varepsilon_1^{LOW} + \varepsilon_2^{HIGH} = A^{SEVERE} = \varepsilon_1^{HIGH} + \varepsilon_2^{LOW}$. 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 - $EU_g = \varepsilon_2 U(w+b) + (1-\varepsilon_2)U(w+b-s)$ - against the expected returns of abstaining from crime - $EU_i = (1-\varepsilon_1)U(w) + \varepsilon_1 U(w-s)$. Deterrence is achieved if $EU_i \geq EU_g$ and thus if

$$\varepsilon_1 \left[U(w) - U(w-s)\right] + \varepsilon_2 \left[U(w+b) - U(w+b-s)\right] \leq U(w) - U(w+b-s).$$

We can see that the left hand side is increasing in both ε_1 and ε_2 , demonstrating that both types of error jeopardize deterrence. The relative impact of the two types of errors depends upon the sign of $[U(w) - U(w-s)] - [U(w+b) - U(w+b-s)]$, and thus the second derivative of the utility function: For risk averse agents (i.e. $EU(.)'' < 0$), a *severe* procedure is less deterring than a lenient procedure; For risk neutral agents (i.e. $EU(.)'' = 0$), *severe* and *lenient* procedures are equally deterring; For risk loving agents (i.e. $EU(.)'' > 0$), a *severe* procedure is more deterring than a *lenient* procedure.

[5]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.

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

Non-monetary sanctions. When the sanction does not have a monetary nature, the results are very similar to those under risk neutrality once we assume separability in the utility arguments (monetary vs. non-monetary payoffs). The utility of the action choices available, staying law-abiding or committing crime, are respectively the following: $EU_i = (1 - \varepsilon_1)U(w) + \varepsilon_1 [U(w) - s]$ and $EU_g = \varepsilon_2 U(w + b) + (1 - \varepsilon_2) [U(w + b) - s]$. Deterrence is obtained for $U(w+b) - U(w) \leq (1 - \varepsilon_1 - \varepsilon_2)s$. Note that for risk neutral subjects, this inequality is identical to that which determines deterrence under monetary sanctions. Based on these results, we state our main hypotheses:

1 *Levels of stealing are identical under severe and lenient procedures.*

There are two reasons why this hypothesis may be rejected. Firstly, under monetary sanctions, risk-averse (risk-loving) subjects would lead to *severe* (*lenient*) procedures being less effective. Secondly, by signalling more strongly that stealing is socially unacceptable (see section 2), *severe* procedures may reduce stealing more.

2 *Levels of stealing are identical under monetary and non-monetary sanctions.*

Again, this hypothesis may fail as a result of either non-risk neutrality or the role of sanctions in communicating the social unacceptability of an action. Recall that our theory predicts that behaviour under the two types of sanctions is identical only for risk neutral subjects. Also, non-monetary sanctions may send a stronger message of social condemnation than fines (see section 2), which could make them more effective in reducing stealing.

4 Experimental Design

The experiment consisted of four parts: i) A slider task; ii) a procedure to elicit subject's willingness to pay (*wtp*) to avoid repeating the slider task; iii) three decisions about how much to steal from another subject; iv) and an elicitation of risk preferences. These incentivised tasks were followed by an unincentivised 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 as a result of the *wtp* elicitation and/or the realized stealing decision were required to do so before collecting their final payment. No subject refused to repeat the tasks when required to do so.

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. A total of 70 subjects participated in the MS treatment, and 76 in the N-MS treatment. The order in which the different error structures were introduced were randomized within each session to control for possible order effects.

We now describe each part of the experiment in detail. Screenshots and full instructions including the text of the trolley dilemma and domain-specific risk questions can be found in the Appendix.

Slider Task. The first part of the experiment was the "slider task" (see Gill and Prowse, 2012 for further details). Subjects were paid 5 Euro 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 (BDM) mechanism to elicit subjects' evaluation of the effort task. Subjects were given an extra 6 Euro 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 which varied between 0 and 6 Euro 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 Euro 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 about how much to steal from another randomly selected subject. They were informed that they were able to take up to 5 Euros from another subject, then were asked "Do you want to take from the other participant? If yes, how much do you want to take?"

In the first decision ($No-sanction$) of all subjects, there was no possibility of punishment. As all treatments were identical up to and including this decision, this gave us a reliable control for social and moral concerns relating to stealing that were unrelated to fear of punishment.

In the second decision, half of the subjects in each of the MS and N-MS treatments faced the *Severe* error structure, and half the *Lenient* error structure. Subjects were informed that after they took 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 could 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. After this second decision, a third decision was made under the alternative error structure.

In the N-MS treatment, the punishment was to place a further 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 in preference to repeating the slider task, and 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.[7]

At the end of the experiment, half of the subjects were randomly selected to have one of their three taking choices implemented (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 (the safe lottery) won either €1.80 or €2.25 with equal probability; those who chose lottery B (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.[8] 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 fill in a questionnaire with the Cognitive Reflection Test (Frederick, 2005) to measure individuals' specific cognitive ability; a selection of questions form 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

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

[8]We chose this method as it has been shown to give the most reliable results of list-based risk elicitation methods (Csermely and Rabas, 2015).

April and October 2015. Subjects were invited using ORSEE (Greiner, 2004). Sessions lasted approximately one hour, and subjects earned between €6.25 and €25.

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 make some attempt to shed light on the mechanisms underlying our treatment effects.

5.1 Overview of Data

For simple treatment comparisons of the level of deterrence 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.[9] The two distributions of the responses to the elicitation procedure are shown in Figure 1 and indeed do not differ significantly between treatments (WMW test: $p = 0.428$).[10]

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 a substantial, albeit not statistically significant, difference between levels of stealing without sanctions between the MS and N-MS treatments (66% and 75% respectively; z-test $p = 0.234$). We remind the reader that both treatments were identical up to and including this decision so the difference is 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, however the difference is only statistically significant for the N-MS treatments. We only find evidence that introducing sanctions changes the distri-

[9]We find no evidence of order effects for any of the four treatments in the proportion of subjects stealing or amount stolen, so have pooled the data. Details of the statistical tests can be found in the Appendix.

[10]In the *wtp* elicitation, a small number of subjects chose to place additional 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 and so assume that their *wtp* is between 0 and €0.5. All results are robust to dropping these subjects.

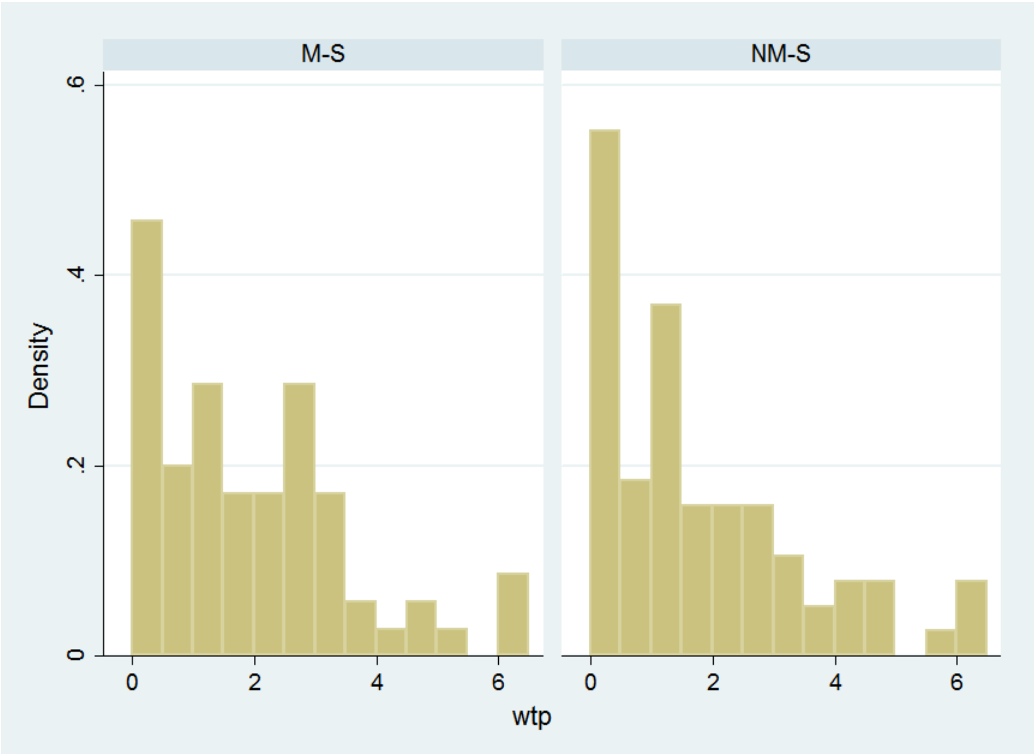


Figure 1: Distribution of *wtp* by treatment.

	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***

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

butions of amounts stolen for the N-MS *Severe* treatment. However, a sign test fails to reject 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 but rather 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 figure 2 which shows the kernel density functions of the amounts taken for the three decisions: under sanctions, either individuals do not steal or they steal everything.

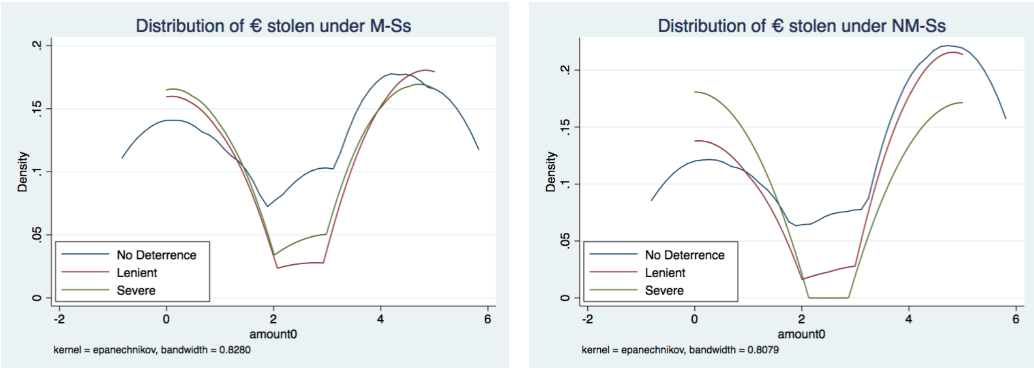


Figure 2: Distributions of amount taken under the three procedures

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 can be seen from table 1, the smallest proportion of subjects steal in N-MS *Severe* (49%) and the greatest in N-MS *Lenient* (63%). Behaviour in the monetary sanction treatments lie in 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 (McNemar test $p = 0.019$).

To control in a simple way for the differences in baseline levels of stealing, we define the dummy variable *deterred* which takes the value one if a subject stole without sanctions, but chose not to when threatened with punishment. We also define the counterpart *encouraged* which takes the value one 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 Figure 3. N-MS *Severe* deterred 30% of individuals, significantly more than the other three treatments for which the figures were 16-17% (MS

Lenient: z-test $p = 0.042$; MS *Severe*: z-test $p = 0.068$; N-MS *Lenient*: McNemar $p = 0.021$). 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%.

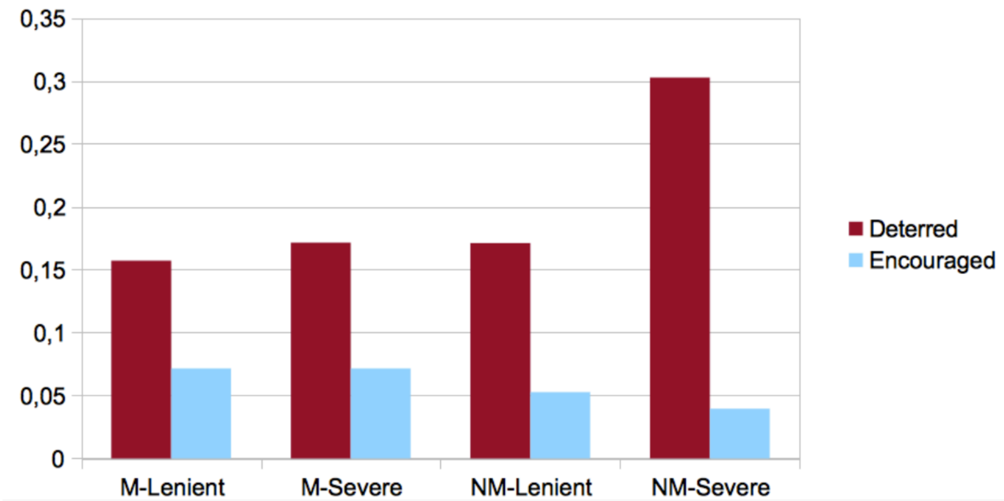


Figure 3: Proportion of subjects stealing in each treatment.

To back up 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). The first model includes only treatment dummies (MS *Severe* 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 or not 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 *Severe* ($p = 0.073$) and N-MS *Lenient* ($p = 0.010$).

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.[11]

Controlling for gender, age, risk-aversion, and cognitive ability makes little difference to the results.

VARIABLES	(1) steal	(2) steal	(3) steal	(4) steal
M_Severe	-0.0143	-0.0181	-0.0155	-0.0484
	(0.0557)	(0.0708)	(0.0747)	(0.0841)
NM_Lenient	0.0610	0.0139	0.00491	0.0260
	(0.0818)	(0.0896)	(0.0909)	(0.141)
NM_Severe	-0.0843	-0.160*	-0.177*	-0.144
	(0.0829)	(0.0899)	(0.0905)	(0.143)
steal1		0.527***	0.544***	0.545***
		(0.0647)	(0.0636)	(0.0627)
sanctions			-0.0680***	-0.0641
			(0.0262)	(0.0447)
sanctions_M_Severe				0.0159
				(0.0541)
sanctions_NM_Severe				-0.0181
				(0.0603)
sanctions_NM_Lenient				-0.0110
				(0.0583)
Observations	292	292	292	292

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Probit models

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, and either *lenient* or *severe* procedures with monetary sanctions.

[11] We comment on Model 4 in section 5.4.

5.3 Treatment comparisons: Amount stolen

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: $p = 0.073$).[12] 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 compare the distributions of the differences between the amount a subject stole without sanctions and what they stole with sanctions. Doing so shows that the distribution of the changes in stealing in N-MS *Severe* (average=-0.67) is significantly different than in both MS *Lenient* (average=-0.04; MWW $p = 0.070$) and MS *Severe* (average=-0.19; MWW $p = 0.098$). 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[13] find no differences between any treatments at conventional levels of 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.[14] In both regressions the only statistical differences between coefficients suggest that people steal less in N-MS *Severe* than N-MS *Lenient*, significant in both regressions (column (1): $p = 0.031$; column (2): $p = 0.031$).

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 dif-

[12] Details of tests for all treatment comparisons can be found in Appendix 6.

[13] 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. For this reason, when we find a signficant difference using a Mann-Whitney test, we perform in addition a stochastic inequality test (Schlag, 2008) 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.

[14] Controlling also for the size of sanctions makes no difference to any of the results in this section.

VARIABLES	(1) *amount*	(2) *amount*	(3) $amount\|steal$	(4) $amount\|steal$
M_Severe	-0.150	-0.150	-0.151	-0.135
	(0.261)	(0.261)	(0.165)	(0.163)
NM_Lenient	0.371	0.110	0.148	0.0687
	(0.397)	(0.323)	(0.178)	(0.156)
NM_Severe	-0.202	-0.462	0.387**	0.312**
	(0.408)	(0.355)	(0.151)	(0.135)
amount1		0.605***		0.183***
		(0.0646)		(0.0579)
Constant	2.636***	1.017***	4.612***	3.948***
	(0.288)	(0.276)	(0.151)	(0.308)
Observations	292	292	164	164
R-squared	0.009	0.296	0.062	0.216

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Amount stolen

ferences in distributions between N-MS *Severe* and each of the other three treatments, each significant at the 5% level or higher (see table 8).[15] Stochastic inequality tests show that these differences are all 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 (see table 9). 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 third and fourth columns of Table 3 find that, conditional on stealing, subjects in N-MS *Severe* steal more than in each of the other treatments, whether or not the amount stolen without sanctions is controlled for. All these differences are significant at the 5% level or higher.

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.

[15] Here we use only unmatched data tests, as otherwise we can only use data from subjects who stole in both treatments, substantially reducing our sample size.

b. There is no evidence of a difference in the distributions of amounts stolen between a *lenient* procedure with non-monetary sanctions, and either *lenient* or *severe* procedures with monetary sanctions.

Result 2: 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, and either *lenient* or *severe* procedures with monetary sanctions.

5.4 Further Results

So far we have rejected the hypotheses that neither error structure nor sanction type affect the level of stealing: N-MS *Severe* reduces stealing at both the intensive and extensive margin more than either N-MS *Lenient* or either of the MS treatments. 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 two regimes. In this section we try to shed some light as to 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 deterred 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.

On the other hand, the direction of treatment differences are almost entirely consistent with the hypothesised effect of changes in the strength of

	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

Table 4: Proportion of subjects deterred.

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 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 look to see if the marginal effect of sanctions differ 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 ($p = 0.962$). In addition, 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 sanction, 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 Conclusion

A major trend in contemporary deterrence policies, at least outside the United States, is to substitute non-monetary sanctions with monetary ones. This trend can be traced back to the major impact that Becker's theory of optimal deterrence had on policy making. The theory suggests to substitute N-MSs with MSs, once the monetary equivalent of the disutility of the punishment is adequately computed for. However, this implication rests on the implicit assumption that MSs and N-MSs are perfect substitutes. In this paper we have tested experimentally this assumption and our results suggest

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

that this substitutability assumption needs to be considered with more caution. Of course one might be skeptical of the external validity of any finding in the lab concerning criminal deterrence policy but our ambition is simply to falsify the optimal deterrence theory. Rather than suggesting any specific policy implication we are here providing a robust test of a theory that has already produced a major impact on policy.

In the experiment we have elicited the willingness to pay (*wtp*) to avoid the effort task for every subject. This same task was later used in our hard labour treatment as a N-MS: every convicted subject had to carry out the effort task before leaving the lab. By eliciting the *wtp* we know what is the monetary equivalent of the N-MS we impose on each individual. Furthermore, in the monetary treatment we impose a sanction equal to their *wtp* to avoid the task.

Our results show that with a *severe* adjudication procedure, N-MSs reduce subjects' propensity to steal relative to MSs. This suggests that when imposing N-MSs instead of MSs with the same "monetary equivalent", the authority convey a whole different message to subjects, a message that is able to induce more of them to switch behavior from crime to no-crime. In order to understand why, perhaps we should look beyond deterrence. Law & Economics scholars have long recognized that the law also has an expressive function in as much as it offers guidance on what is the appropriate behavior the society expects individuals to follow (see Sunstein 1996; Nance 1997; Cooter 1998; Funk 2007 for a sample of the literature). On the other hand D'Antoni and Galbiati (2007) noted that the potential use of MSs as a mean 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.

With respect to comparisons between *severe* and *lenient* adjudication 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. (2014), 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 finding, and provide more conclusive evidence on why theoretically equivalent sanction regimes may or may not be more effective in reducing crime.

Appendix: Order Effects

Treatment	First Choice	Second Choice	MW 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	0.50
M-S Severe	2.54	2.44	0.89

Figure 4: Order Effects: Amount Stolen

Treatment	First Choice	Second Choice	z-test p-value
N-MS Lenient	0.58	0.68	0.342
N-MS Severe	0.45	0.53	0.491
M-S Lenient	0. 54	0.60	0.629
M-S Severe	0. 57	0.54	0.810

Figure 5: Order Effects: Proportion Stealing

Appendix: Treatment Effects

	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*	

Figure 6: 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	-0.06	0.09		n.s.
N-MS Severe	-0.63*	-0.48*	-0.57*	

Figure 7: Change in 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)

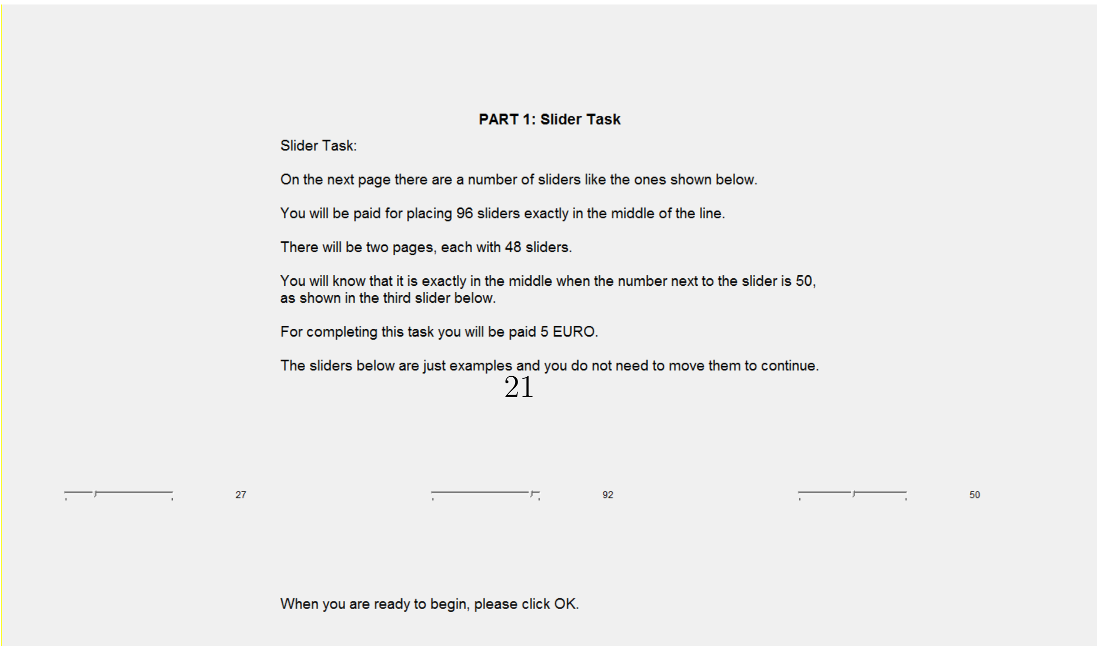
	M-S Lenient	M-S Severe	N-MS Lenient	N-MS Severe
M-S Lenient (n=39)		n.s.	n.s.	**
M-S Severe (n=40)	-0.15		n.s.	***
N-MS Lenient (n=48)	0.15	0.1035		**
N-MS Severe (n=37)	0.39***	0.54***	0.24**	

Figure 8: Amount stolen conditional on stealing - Lower left: comparisons of distributions (WMW); Upper right: comparisons of stochastic inequality (Stochastic inequality test)

	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)	-0.06		n.s.	n.s.
N-MS Lenient (n=48)	-0.28	-0.22		n.s.
N-MS Severe (n=37)	-0.02	0.04	0.26	

Figure 9: Amount stolen conditional on stealing - Lower left: comparisons of distributions (WMW); Upper right: comparisons of stochastic inequality (Stochastic inequality test)

Appendix: Instructions & Screenshots



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

Figure 11: **Instructions on the *wtp* to avoid the task elicitation phase**

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

Figure 12: **General Instructions on stealing phase**

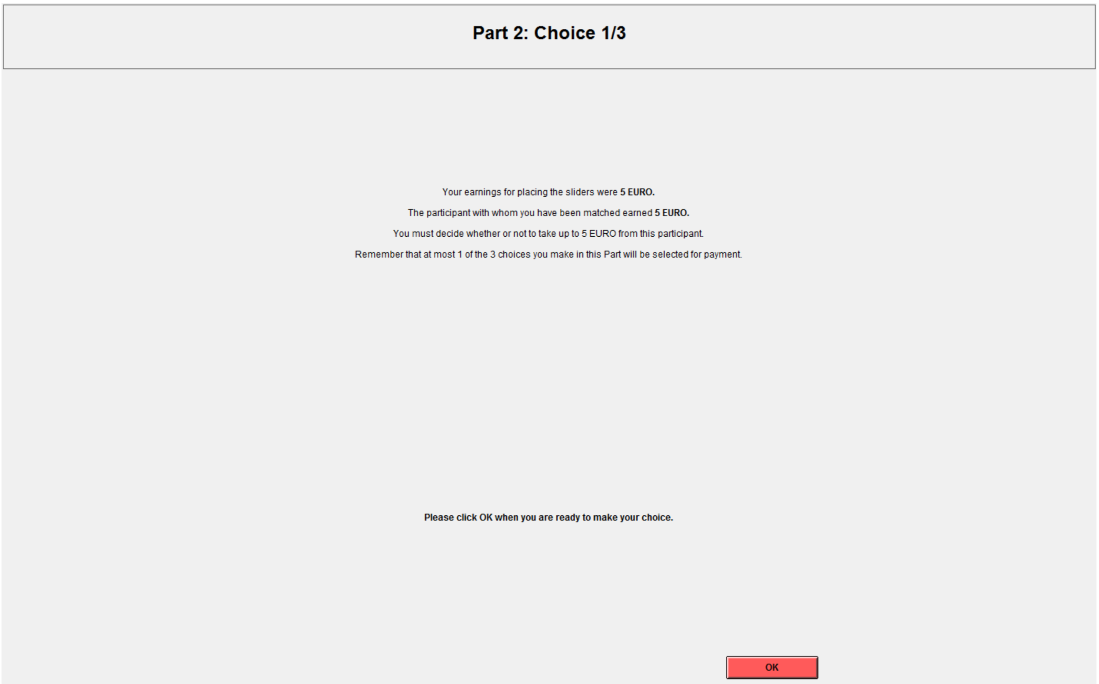
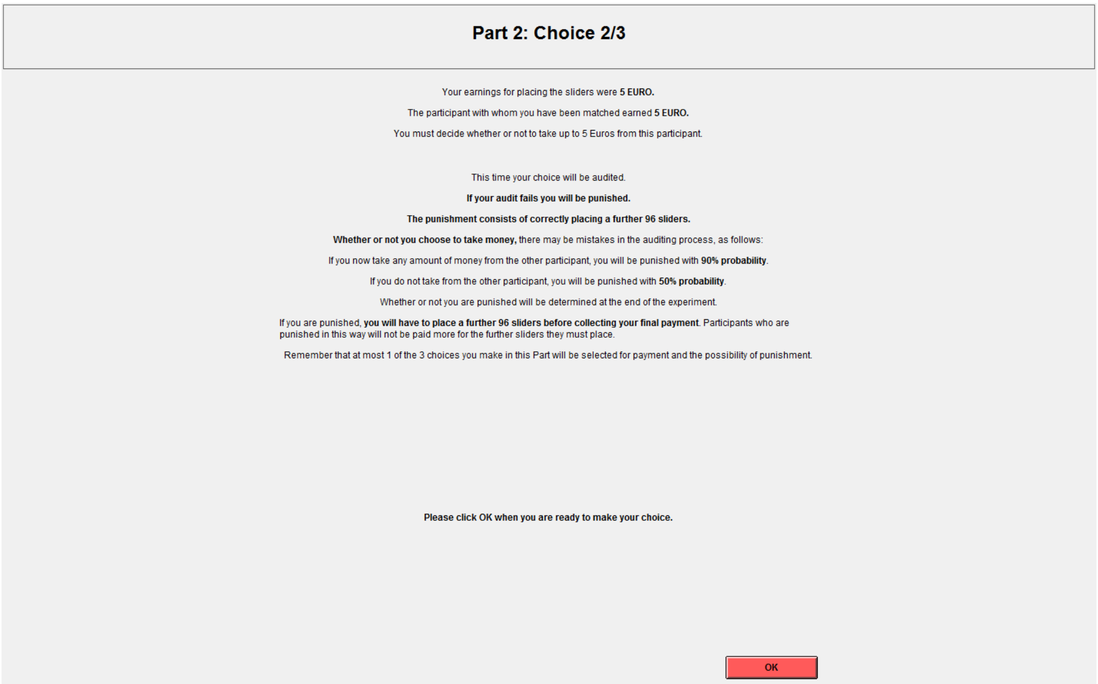


Figure 13: **Instructions on the decision to take in the** $No-deterrence$ **procedure. These instructions were the same for both MSs and N-MSs treatments**



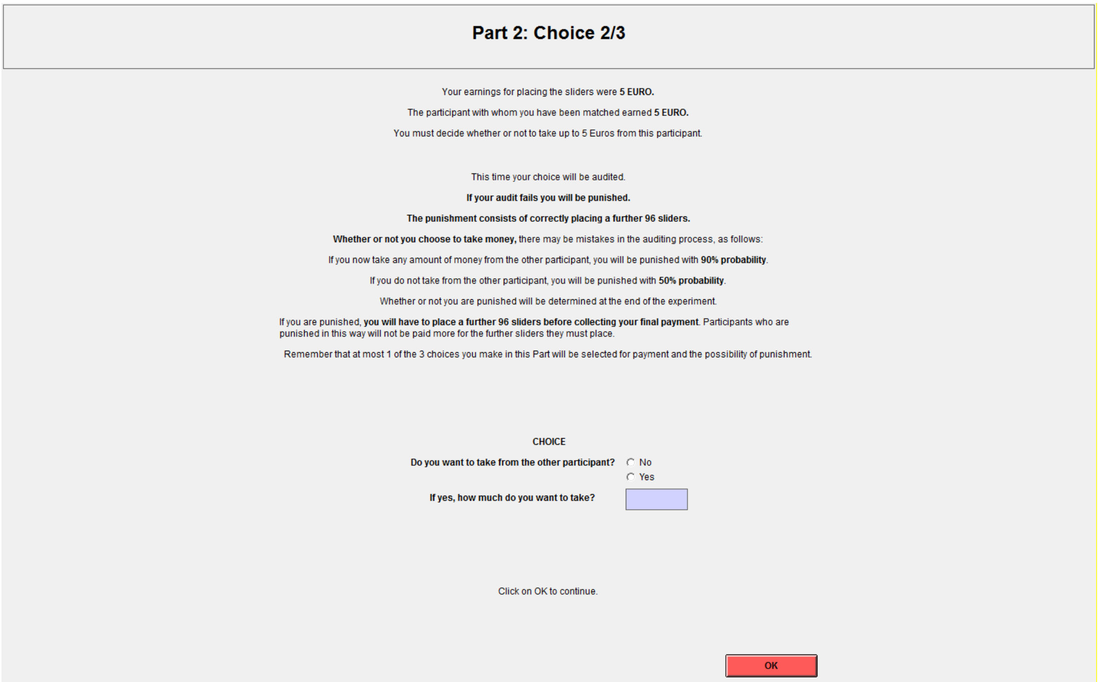
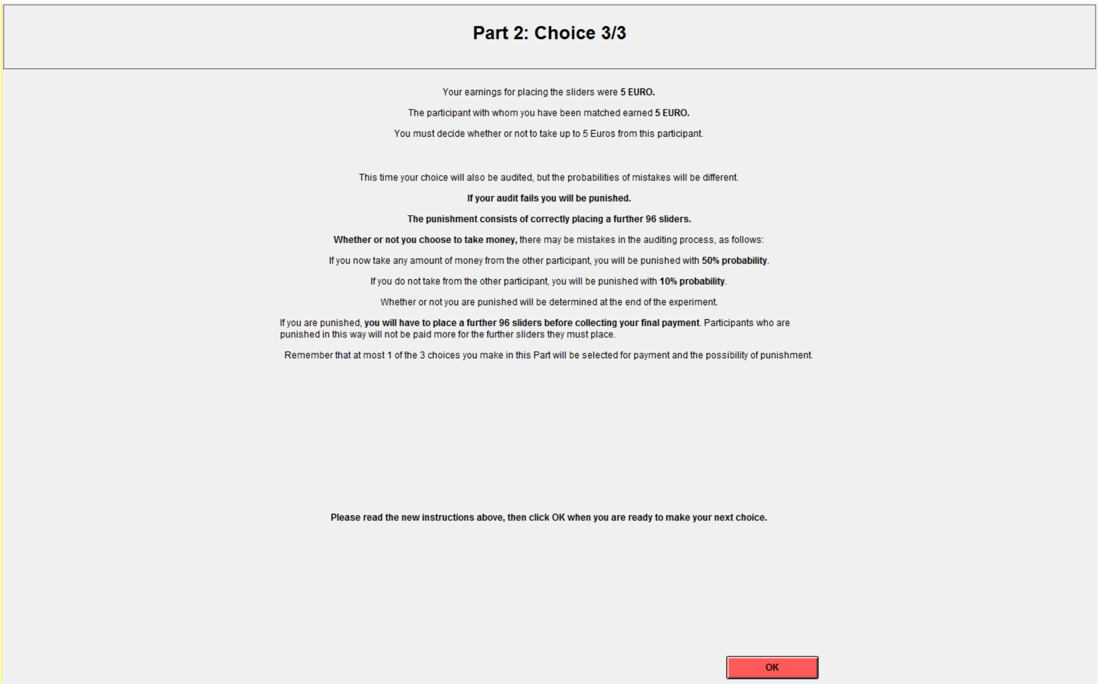


Figure 14: **Instructions on the decision to take in the *Lenient* procedure. Above there are the instructions for the MSs treatment and below the instructions for the N-MSs treatment**



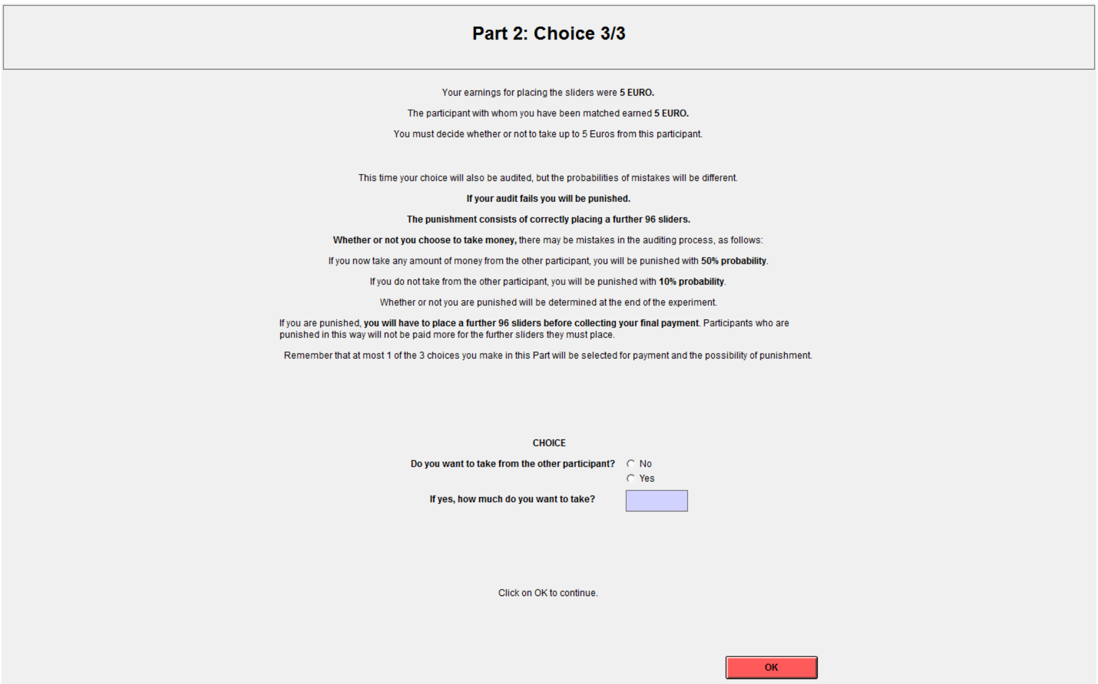


Figure 15: **Instructions on the decision to take in the *Severe* procedure. Above there are the instructions for the MSs treatment and below the instructions for the N-MSs treatment**

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

Figure 16: Instructions on the risk elicitation mechanism à-la Holt and Laury (2002)

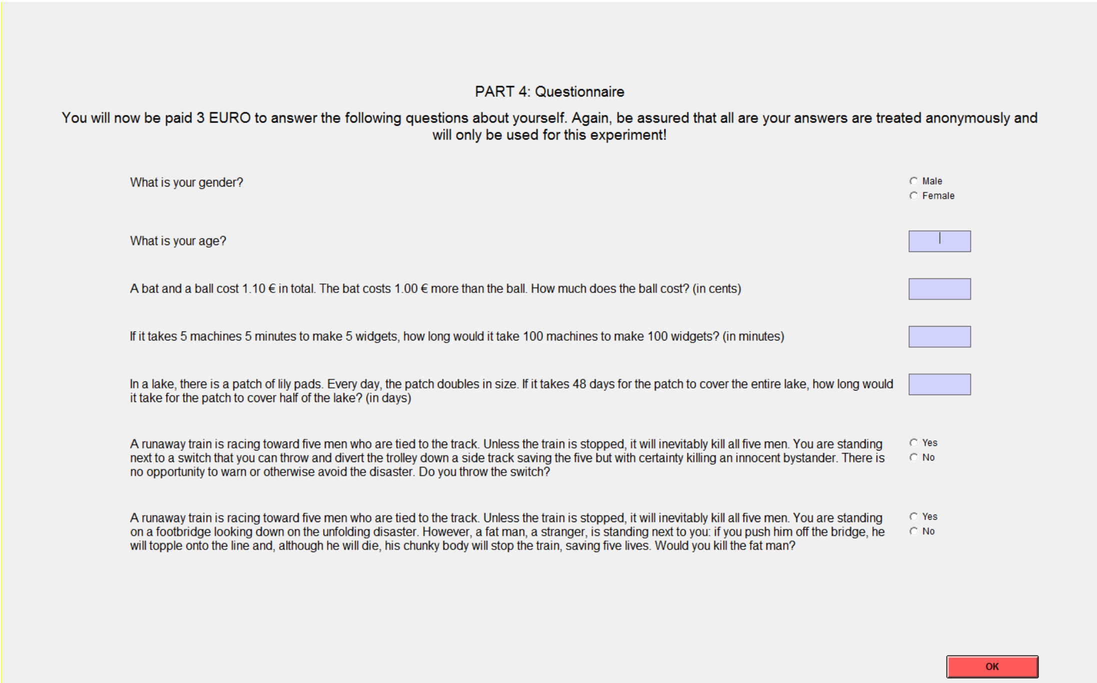
PART 4: Questionnaire

For each of the following statements, please indicate how likely it is that you would engage in the described activity or behaviour if you were to find yourself in that situation:

Admitting that your tastes are different from a friend.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Investing 10% of your annual income in Greek government bonds.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Taking some questionable deductions on your income tax return.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Disagreeing with an authority figure on a major issue.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Betting a days income at a high-stake poker game.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Having an affair with a married man/woman.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Claimng someone else's work is your own.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Investing 5% of your annual income in a very speculative stock.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Betting a day's income on a sporting event.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Revealing a friend's secret to someone else.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Driving a car without wearing a seat belt.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Investing 10% of your income in a new business venture.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Choosing a career that you truly enjoy over a more secure one.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Speaking your mind about an unpopular issue in a meeting at work.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Leaving your young children alone at home while running an errand.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely
Not returning a wallet you found that contains €200.	Extremely Unlikely	○ ○ ○ ○ ○ ○ ○	Extremely Likely

OK

Figure 17: **Instructions on the domain specific risk elicitation mechanism à-la Blais and Weber (2006)**



PART 4: Questionnaire

You will now be paid 3 EURO to answer the following questions about yourself. Again, be assured that all are your answers are treated anonymously and will only be used for this experiment!

What is your gender?	Male Female
What is your age?	
A bat and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost? (in cents)	
If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (in minutes)	
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (in days)	
A runaway train is racing toward five men who are tied to the track. Unless the train is stopped, it will inevitably kill all five men. You are standing next to a switch that you can throw and divert the trolley down a side track saving the five but with certainty killing an innocent bystander. There is no opportunity to warn or otherwise avoid the disaster. Do you throw the switch?	Yes No
A runaway train is racing toward five men who are tied to the track. Unless the train is stopped, it will inevitably kill all five men. You are standing on a footbridge looking down on the unfolding disaster. However, a fat man, a stranger, is standing next to you: if you push him off the bridge, he will topple onto the line and, although he will die, his chunky body will stop the train, saving five lives. Would you kill the fat man?	Yes No

OK

References

Avio, K. L. (1993). Economic, retributive and contractarian conceptions of punishment. *Law and Philosophy 12*(3), 249–286.

Beccaria, C. (1763). *Dei delitti e delle pene*. F. Sanvito.

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

Beckett, K. and A. Harris (2011). On cash and conviction. *Criminology & Public Policy 10*(3), 509–537.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. London,: Printed for W. Pickering [etc.].

Blais, A.-R. and E. U. Weber (2006). A domain-specific risk-taking (dospert) scale for adult populations. *Judgment and Decision Making 1*(1).

Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature 50*(2), 368–425.

Cinyabuguma, M., T. Page, and L. Putterman (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics 89*(8), 1421–1435.

Cooter, R. (1998). Expressive law and economics. *The Journal of Legal Studies 27*(2), 585–608.

Csermely, T. and A. Rabas (2015). How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods. *Available at SSRN 2573544*.

D'Antoni, M. and R. Galbiati (2007). A signalling theory of non-monetary sanctions. *International Review of Law and Economics 27*(2), 204–218.

Edmonds, D., S. Duncan, and G. Armstrong (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. *The American Economic Review*, 307–322.

Feess, E., M. Schramm, and A. Wohlschlegel (2014). The impact of fine size and uncertainty on punishment and deterrence: Evidence from the laboratory. *SSRN*.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25–42.

Friedman, D. D. (1981). Reflections on optimal punishment, or: should the rich pay higher fines? *Research in Law and Economics 3*, 185–205.

Funk, P. (2007). Is there an expressive function of law? an empirical analysis of voting laws with symbolic fines. *American Law and Economics Review*.

Garoupa, N. and D. Klerman (2004). Corruption and the optimal use of non-monetary sanctions. *International Review of Law and Economics 24*(2), 219–225.

Gibbs, J. P. (1968). Crime, punishment, and deterrence. *The Southwestern Social Science Quarterly*, 515–530.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review 102*(1), 469–503.

Gneezy, U. and A. Rustichini (2000). Fine is a price, a. *J. Legal Stud. 29*, 1.

Grechenig, K., A. Nicklisch, and C. Thöni (2010). Punishment despite reasonable doubt—a public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies 7*(4), 847–867.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working paper series in economics 10*(23), 63–104.

Harbaugh, W. T., N. H. Mocan, and M. S. Visser (2011). Theft and deterrence. *National Bureau of Economic Research Working paper 17059*.

Harris, A., H. Evans, and K. Beckett (2010). Drawing blood from stones: Legal debt and social inequality in the contemporary united states1. *American Journal of Sociology 115*(6), 1753–1799.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *The American Economic Review 92*(5), 1644–1655.

Kahan, D. M. (1996). What do alternative sanctions mean? *The University of Chicago Law Review 63*(2), 591–653.

Kahan, D. M. (2005-2006). What's really wrong with shaming sanctions. *Texas Law Review 84*, 2075–2096.

Kahan, D. M. and E. A. Posner (1999). Shaming white-collar criminals: A proposal for reform of the federal sentencing guidelines. *Journal of Law and Economics 42*(1), 365–391.

Kan, S. S. (1996). Corporal punishments and optimal incapacitation. *The Journal of Legal Studies 25*(1), 121–130.

Khadjavi, M. (2015). On the interaction of deterrence and emotions. *Journal of Law, Economics, and Organization 31*(2), 287–319.

Levitt, S. D. (1997a). Incentive compatibility constraints as an explanation for the use of prison sentences instead of fines. *International Review of Law and Economics 17*(2), 179–192.

Levitt, S. D. (1997b). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review 87*(3), 270–290.

Marchegiani, L., T. Reggiani, and M. Rizzolli (2013). Severity vs. leniency bias in performance appraisal: Experimental evidence. Technical report, School of Economics and Management at the Free University of Bozen.

Markel, D. and C. Flanders (2010). Bentham on stilts: The bare relevance of subjectivity to retributive justice. *California Law Review 98*, 907–988.

Markussen, T., L. Putterman, and J.-R. Tyran (2014). Judicial error and cooperation. Technical report.

Masclet, D., C. Noussair, S. Tucker, and M. Villeval (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 366–380.

McAdams, R. H. (2000). An attitudinal theory of expressive law. *Oregon Law Review 79*, 339–390.

Miceli, T. J. (2010). A model of criminal sanctions that incorporate both deterrence and incapacitation. *Economics Letters 107*(2), 205–207.

Montag, J. and T. Sobek (2014). Should paris hilton receive a lighter prison sentence because she's rich? an experimental study. *Kentucky Law Journal, Forthcoming*.

Montag, J. and J. Tremewan (2016). Let the punishment fit the criminal: an experimental study.

Mungan, M. C. (2012). The law and economics of fluctuating criminal tendencies and incapacitation. *Maryland Law Review*.

Murphy, K. R. and W. K. Balzer (1989). Rater errors and rating accuracy. *Journal of Applied Psychology 74*(4), 619 – 624.

Nance, D. A. (1997). Guidance rules and enforcement rules: A better view of the cathedral. *Virginia Law Review 83*(5), 837–937.

Noussair, C. and S. Tucker (2005). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry 43*(3), 649–660.

Paternoster, R. (2010). How much do we really know about criminal deterrence? *The Journal of Criminal Law and Criminology*, 765–824.

Peeters, R. and M. Vorsatz (2013). Immaterial rewards and sanctions in a voluntary contribution experiment. *Economic Inquiry 51*(2), 1442–1456.

Pérez, R. L. and H. J. Kiss (2012). Do people accurately anticipate sanctions? *Southern Economic Journal 79*(2), 300–321.

Polinsky, A. M. and S. Shavell (1984a). The optimal use of fines and imprisonment. *Journal of Public Economics 24*(1), 89–99.

Polinsky, A. M. and S. Shavell (1984b). The optimal use of fines and imprisonment. *Journal of Public Economics 24*(1), 89–99.

Polinsky, A. M. and S. Shavell (1992). Enforcement costs and the optimal magnitude and probability of fines. *Journal of Law and Economics*, 133–148.

Posner, R. A. (1980). Retribution and related concepts of punishment. *The Journal of Legal Studies*, 71–92.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), pp. 7–63.

Rizzolli, M. (2016). *Encyclopedia of Law and Economics*, Chapter Adjudication: Type-I and Type-II errors. Springer.

Rizzolli, M. and L. Stanca (2012). Judicial errors and crime deterrence: theory and experimental evidence. *Journal of Law and Economics 55*(2), 311–338.

Ruback, R. B. (2015). Benefits and costs of economic sanctions: Considering the victim, the offender, and society, the. *Minnesota Law Review 99*, 1779.

Schildberg-Hörisch, H. and C. Strassmair (2012). An experimental test of the deterrence hypothesis. *Journal of Law, Economics, and Organization 28*(3), 447–459.

Schlag, K. H. (2008). A new method for constructing exact tests without making any assumptions. *Available at SSRN 1285454*.

Shavell, S. (1986). The judgment proof problem. *International Review of Law and Economics 6*(1), 45–58.

Shavell, S. (2015). A simple model of optimal deterrence and incapacitation. *International Review of Law and Economics 42*, 13–19.

Sunstein, C. R. (1996). On the expressive function of law. *University of Pennsylvania Law Review 144*(5), pp. 2021–2053.

Travis, J., B. Western, S. Redburn, et al. (2014). *The Growth of Incarceration in the United States:: Exploring Causes and Consequences*. National Academies Press.

Wittman, D. (1974). Punishment as retribution. *Theory and Decision 4*(3-4), 209–237.