UNIVERSITA' DEGLI STUDI DI MILANO
AUDENTES FORTUNA IUVAT
BICOCCA

DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

Judicial Errors and Crime Deterrence: Theory and Experimental Evidence

Matteo Rizzolli and Luca Stanca

No. 170 – August 2009

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
http://dipeco.economia.unimib.it

Judicial Errors and Crime Deterrence: Theory and Experimental Evidence

Matteo Rizzolli[*] and Luca Stanca[†]

August 2009

Abstract

The standard economic theory of crime deterrence predicts that the conviction of an innocent (type-I error) is as detrimental to deterrence as the acquittal of a guilty individual (type-II error). In this paper, we qualify this result theoretically, showing that in the presence of risk aversion, loss-aversion, or differential sensitivity to procedural fairness, type-I errors can have a larger effect on deterrence than type-II errors. We test these predictions with an experiment where participants make a decision on whether to steal from other individuals, being subject to different probabilities of judicial errors. The results indicate that both types of judicial errors have a large and significant impact on deterrence, but these effects are not symmetric. An increase in the probability of type-I errors has a larger negative impact on deterrence than an equivalent increase in the probability of type-II errors. This asymmetry is largely explained by risk aversion and, to a lesser extent, type-I error aversion.

Keywords: Judicial errors, criminal procedure, procedural fairness, experimental economics, law and economics, crime, deterrence.
JEL codes: C91, K14, K41, K42.

[*] Economics Department, University of Milan - Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: matteo.rizzolli@unimib.it

[†] Corresponding author. Economics Department, University of Milan - Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: luca.stanca@unimib.it

> *"Un coupable puni est un exemple pour la canaille; un innocent condamné est l'affaire de tous les honnêtes gens."*[1]
>
> (Jean de La Bruyère, Les Caractères, 1692)

> *"The prospect of innocents languishing in prison or, worse, being put to death for crimes that they did not commit, should be intolerable to every American, regardless of race, politics, sex, origin, or creed."*
>
> (The Innocence Project, 2008)[2]

1 Introduction

It is common wisdom to consider the punishment of an innocent more questionable than the acquittal of a guilty individual. As *honnêtes gens* – in the words of de La Bruyère – we share the same opinion. As economists, however, we must recognize that the support for this view is less clear-cut. In fact, the standard economic theory of public enforcement of law suggests that the conviction of an innocent (type-I error) is no worse than the acquittal of a guilty individual (type-II error), since both types of errors jeopardize deterrence by the same token (see for instance Polinsky and Shavell, 2007). The implication of this results is straightforward: if optimal deterrence is what matters, the policy-maker – and the judge – should be indifferent between one additional type-I error and one additional type-II error.

The economic model of crime deterrence, originally developed by Becker (1968), was extended by Harris (1970) to include the effect of type-I errors. Png (1986) modelled explicitly the effect of both types of judicial errors on deterrence: a higher probability of type-I errors decreases the expected payoff of abiding by the law, whereas a higher probability of type-II errors increases the payoff of engaging in the unlawful activity. Thereafter, this extension has been generally incorporated in the economic literature on crime deterrence (see e.g. Kaplow, 1994; Garoupa, 1997; Polinsky and Shavell, 2008). Within this framework, the effects of judicial errors on deterrence are expected to be symmetric, since the probabilities of type-I and type-II errors have the same impact on the difference between the returns from honesty and from crime.

[1] *A guilty man punished is an example for the rabble; an innocent man condemned is a matter for all honest people.*

[2] The Innocence Project is a non-profit, national litigation and public policy organization dedicated to exonerating wrongfully convicted people through DNA testing and reforming the criminal justice system to prevent future injustice. Through the years the project has managed to obtain the exoneration of 225 people that received a final sentence; 17 of which were on the death row. The quotation is from the Innocence Project website, http://www.innocenceproject.org.

This theoretical account of the relationship between judicial errors and crime deterrence has been the object of several criticisms. Ehrlich (1982) observes that the conviction of an innocent may increase deterrence if it is perceived by other imperfectly informed would-be offenders as the conviction of a guilty individual. They may misinterpret an increase of type-I errors as a decrease of type-II errors, and can therefore be more strongly deterred from committing crimes. Strandburg (2003) notices how the two types of judicial errors are inextricably linked together, as the production of both depends on the same enforcement strategy by the authority. Depending on the strategy implemented, an increase in accuracy, represented by a reduction of the sum of the two errors (Kaplow, 1996), may result in either higher or lower deterrence.[3] Kaplow and Shavell (1994) point out how judicial errors generally discourage participation in socially desirable activities. This effect becomes stronger when risk-aversion is also considered (Block and Sidak, 1980).[4] More recently, the Png (1986) indifference result has been questioned by Lando (2006), who argues that when mistakes about the identity of the criminal are taken into account,[5] type-I errors have no detrimental effect on deterrence (see Garoupa and Rizzolli, 2009, for a critique of Lando's argument). Fon and Schaefer (2007) show that the negative effect of type-I errors on deterrence can be partially offset by state liability against wrongful convictions.

At the empirical level, several experimental analyses have examined the deterrence hypothesis,[6] although only relatively few are based on a setting that convincingly reproduces crime in the lab. Falk and Fischbacher (2002) examine how social interaction affects the propensity to commit a crime. Visser et al. (2006) explore the deterrence hypothesis experimentally using a reverse dictator game. Subjects are randomly paired and must decide how much to steal from their counterpart's endowment, having been informed of the exogenously determined probability of type-II error. By varying both the potential amount to be stolen and the probability of detection, the authors study whether the decisions to steal satisfy the Generalized Axiom of Revealed Preference (GARP).[7] Hoerisch and Strassmair (2008) conduct an

[3] The nuisance effect of errors on the perception of the rule in place has also been noticed by Craswell and Calfee (1986).

[4] See Immordino and Polo (2008) for an analysis of the effect of judicial errors on the innovative activity of firms.

[5] Mistakes of identity are those for which, in the presence of evident crimes such as murders and robberies, the wrong person is incriminated. Mistakes of act occur instead when someone is convicted for crimes that did not happen.

[6] Monitoring and sanctioning scofflaw behavior is routinely part of experiments in personnel economics, such as Backes-Gellner et al. (2008) and Falk and Gächter (2008). See also Schulze and Frank (2003); Abbink (2006) on the analysis of corruption, Torgler (2002) on tax compliance, and Fehr and Gachter (2000) on public good experiments.

[7] The authors also estimate demand functions for stolen loot and the corresponding elasticities for criminal participation and the amount of money stolen with respect to the terms of trade between money stolen and the probability of detection, and between money

experiment on crime deterrence using a similar reverse dictator game. They examine whether observed decisions to steal are compatible with alternative theories of social preferences, finding that the deterrence hypothesis poorly describes the observed behavior. When there is little or no probability of detection, so that the incentives for theft are high, subjects steal less than when the expected sanction is more significant (however, when the sanction becomes severe, the deterrence hypothesis holds). They conclude that the threat of a sanction may crowd out law-abiding behavior.[8]

Judicial errors are an important issue for public policy. They are particularly relevant, although not exclusively, for the enforcement of administrative and criminal law. Quite surprisingly, however, the economic theory of optimal deterrence has so far paid relatively little attention to the effects of judicial errors. At the empirical level, most of the existing experimental work considers only the case where criminal behavior goes unpunished (type-II error), whereas there is hardly any evidence on the case where law-abiding behavior is erroneously sanctioned (type-I error).

In this paper, we build on the theory of optimal deterrence to show that there are several economic arguments that may explain asymmetric effects of type-I and type-II errors on deterrence. First, the Png (1986) equivalence result is not robust to departures from risk-neutrality. By simply introducing risk-aversion in the standard model of public enforcement, it can be shown that type-I errors are more detrimental to deterrence than type-II errors. Second, the introduction of loss-aversion reinforces this result, so that a given increase in the probability of type-I errors, compensated by an equivalent decrease in the probability of type-II errors, induces more people to commit the crime. Third, we propose a further behavioral element that can explain differences in responses to the two type of errors: a differential sensitivity to procedural fairness that results in what we refer to as *type-I error aversion*: being punished when one has abided by the law implies a specific cost in terms of loss of guidance and motivational crowding-out.

We then provide empirical evidence on these theoretical hypotheses by means of an appropriately designed laboratory experiment. The main objective of the analysis is to test the deterrence hypothesis, focusing on the role of type-I errors, and to assess whether type-I and type-II errors have the same impact on deterrence. Our findings indicate that both types of judicial errors have a large and significant impact on deterrence. However, contrary to the predictions of the standard theory of crime deterrence, type-I and type-II errors do not have symmetric effects. An increase in the probability of type-I errors is found to have a larger negative impact on deterrence than the same increase in the probability of type-II errors. This asymmetry is largely explained by the effect of risk aversion and, to a lesser extent, sensitivity to

stolen and the sanction.

[8] See also Sonnemans and van Dijk (2008) for an analysis of the role of judicial errors from the judge's viewpoint.

procedural fairness, whereas loss aversion does not play a significant role.

The paper is structured as follows. In section 2 we provide the theoretical framework. Building on the standard theory of crime deterrence, we focus on the role of type-I errors and extend the model to account for risk attitudes, loss aversion and sensitivity to procedural fairness. Section 3 describes the experimental design. Section 4 presents the results. Section 5 concludes with a discussion of the main findings and the implications of the analysis.

2 Theory

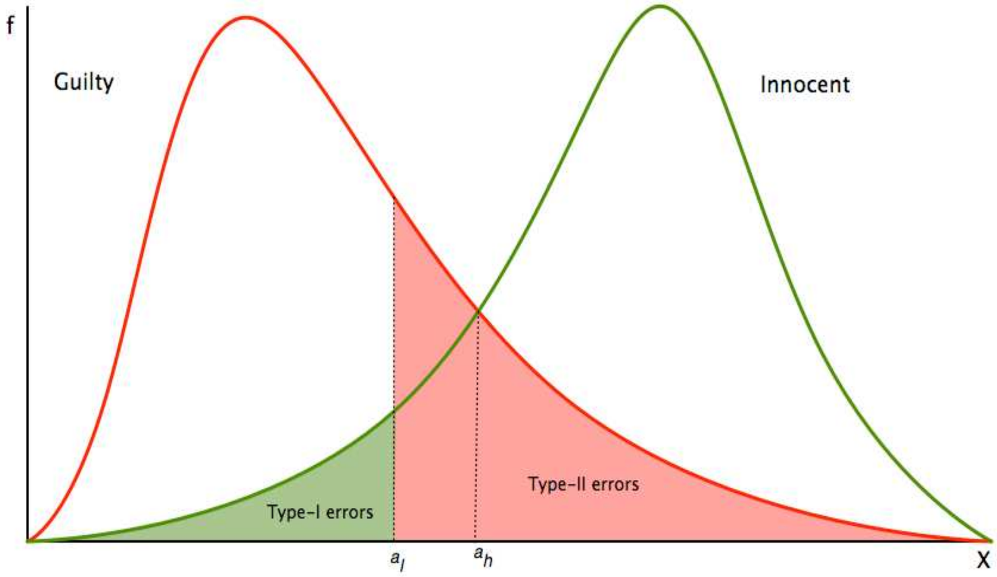
The basic model of crime deterrence envisages an individual that must decide whether to commit a crime. His ex-ante wealth is w_0 and a successful crime makes him gain g. The adjudicative authority monitors the criminal activity and pursues the case if it detects it.[9] The authority must convince the judge that the individual deserves a conviction. The adjudication is prone to two types of judicial errors: innocent individuals mistakenly judged guilty (type-I errors) and guilty individuals mistakenly judged innocent (type-II errors).

Defendants' ability to confute the charges of the prosecutor derives from factors that are either dependent on their actual innocence (such as the ability to produce exculpatory evidence) or independent of it (such as wealth and the possibility to afford good lawyers). The probability of successfully confuting the prosecutor's allegations is a function of the frequency distribution of these differential abilities. There are therefore two distributions for the probability of the accused to successfully defend oneself, one for guilty individuals and one for innocent individuals, as described in Figure 1. The null hypothesis is that the accused subject is innocent. The prosecutor submits his charges to the court seeking to convince the judge to refute the null hypothesis. The defendant confutes the prosecutor's charges in order to support the null hypothesis. Let X be the ability of the defendant to confute the prosecutor's allegations. We assume that, on average, innocent individuals have a high X while guilty individuals have a low X, but the two distributions overlap.

Suppose that, in a given procedure, the required burden of proof to convict is low. This means that the prosecutor can easily convince the judge to reject the null and, conversely, the defendant must have a high ability (a_h) to defend himself if he wants to be acquitted. Notice that, to the left of a_h there will be a fraction of innocent individuals who will not be able to successfully defend themselves and will be wrongfully convicted, so that the probability of type-I error is ε_1. Conversely, to the right of a_h there will be a fraction of guilty individuals who are able to prove their innocence, so that the probability of type-II error is ε_2. If the standard of evidence is increased, for instance, from "preponderance of evidence"to "beyond any

[9] For simplicity, we do not distinguish between the enforcement authority, such as the police, in charge of detecting the crime, and the adjudicative authority itself, such as the court, that decides wether the evidence collected is sufficient to reach a conviction.

Figure 1: Frequency distributions of the authority's ability to prove guilt



Note: The figure is adapted from Feinberg (1971)

reasonable doubt", the ability required to prove one's own innocence falls to a_l. This results in higher ε_2 and lower ε_1.[10]

Let us go back to the choice between committing the crime (C) or being innocent (I). The individual knows that, if he decides to commit the crime, he faces a probability ε_2 of successfully escaping the conviction. Conversely, if he does not commit the crime, he faces a probability ε_1 of being convicted even if innocent. If convicted, he faces a fixed sanction f, that represents the magnitude of the fine or the private costs of the prison term.

First, assume risk neutrality. In order to decide whether to commit the crime, each individual assesses the net benefits of crime (the gains g minus the expected costs of committing the crime) against the expected costs of staying honest (positive because of type-I errors). The expected costs of crime are determined by the magnitude of the fine and/or the costs of the sanction (f), and by the probability of being eventually convicted $(1-\varepsilon_2)$. The costs of staying honest are determined by ε_1 and f. The expected payoffs in the two cases are:

[10] From basic decision theory, it is well known that a decision rule prescribing the rejection of the less probable hypothesis minimizes expected error. Therefore, for a given technology of fact-finding, accuracy (which is measured with the sum of the two errors) is maximized when the standard of proof is set at the "preponderance of evidence" level (Demougin and Fluet, 2005). In Figure 1 this is obtained at a_h. Accuracy can also improve if the technology of fact finding improves. In this way both types of errors simultaneously decrease and thus the judge can better discriminate between innocence and guilt (Kaplow, 1994; Kaplow and Shavell, 1994). In Figure 1 this would result in the two distributions being less dispersed around the mean and thus having a smaller overlapping region.

$$\begin{cases} E\pi_I = \varepsilon_1(w_0 - f) + (1 - \varepsilon_1)w_0 \\ E\pi_C = \varepsilon_2(w_0 + g) + (1 - \varepsilon_2)(w_0 + g - f) \end{cases} \tag{1}$$

The individual will not commit the crime if the expected payoff from the criminal activity does not exceed the expected payoff from innocence, that is if $E\pi_C < E\pi_I$. This is true if

$$\frac{f}{g} > \frac{1}{1 - \varepsilon_1 - \varepsilon_2} \tag{2}$$

Notice from equation (2) that ε_1 and ε_2 have the same impact on deterrence. On the one hand, type-II errors undermine deterrence as they decrease the probability of being convicted for guilty individuals. On the other hand, type-I errors increase the costs of staying honest and thus decrease the relative costs of committing the crime. At the margin, one further innocent convicted is equivalent to one further guilty acquitted (Png, 1986).[11]

Proposition 1. *If individuals are risk-neutral, type-I and type-II errors are equally detrimental to deterrence.*

2.1 Risk aversion

In the presence of risk aversion, type-I errors are more detrimental to deterrence than type-II errors. Assuming a concave utility function, the expected utility from crime and innocence can be written as follows:

$$\begin{cases} EU_I = \varepsilon_1 U(w_0 - f) + (1 - \varepsilon_1)U(w_0) \\ EU_C = \varepsilon_2 U(w_0 + g) + (1 - \varepsilon_2)U(w_0 + g - f) \end{cases} \tag{3}$$

A rational individual will be deterred from committing the crime if $EU_I > EU_C$, that is if

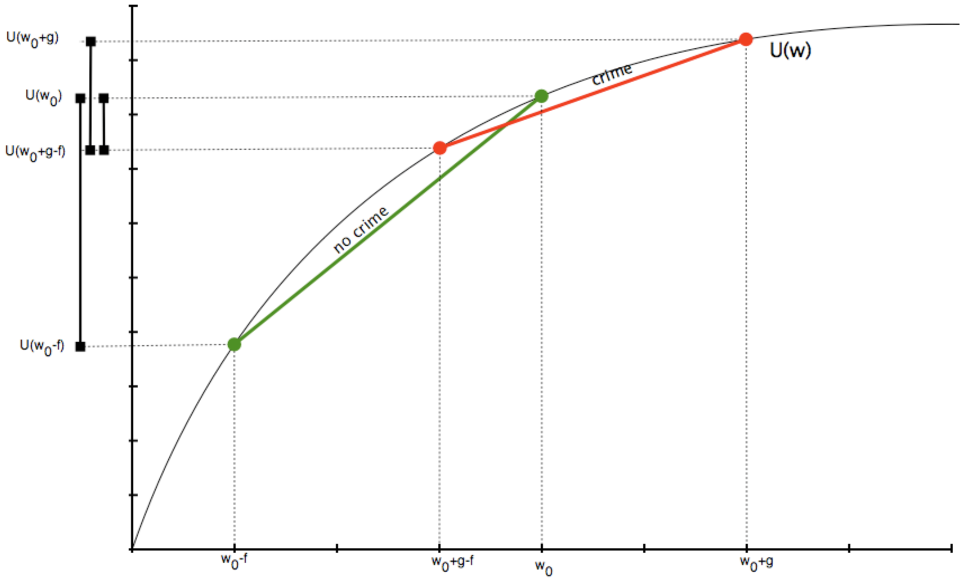
$$\begin{split} U(w_0) - U(w_0 + g - f) - \varepsilon_1[U(w_0) - U(w_0 - f)] \\ - \varepsilon_2[U(w_0 + g) - U(w_0 + g - f)] > 0 \end{split} \tag{4}$$

Figure 2 shows how both ε_1 and ε_2 negatively affect deterrence, as in the case of risk neutrality. However, given the concavity of the utility function, the negative impact of type-I errors on expected utility is larger. To see why, notice that the concavity of the utility function implies that $U(w_0) - U(w_0 - f) > U(w_0 + g) - U(w_0 + g - f)$. Therefore ε_1 has a relatively larger adverse impact on deterrence than ε_2. Put it in a different way, in order to maintain the same level of deterrence, a given increase in ε_1 must be compensated by a larger decrease in ε_2.

[11] The standard analysis of optimal deterrence extended to include type-I errors usually considers only the risk-neutral case (see for instance Polinsky and Shavell (2007), section 15).

Proposition 2. *In the presence of risk aversion, type-I errors are more detrimental to deterrence than type-II errors*

Figure 2: Expected utility for a risk-averse individual



Note: The figure displays the expected utilities EU_I and EU_C as a function of ε_1 and ε_2, respectively.

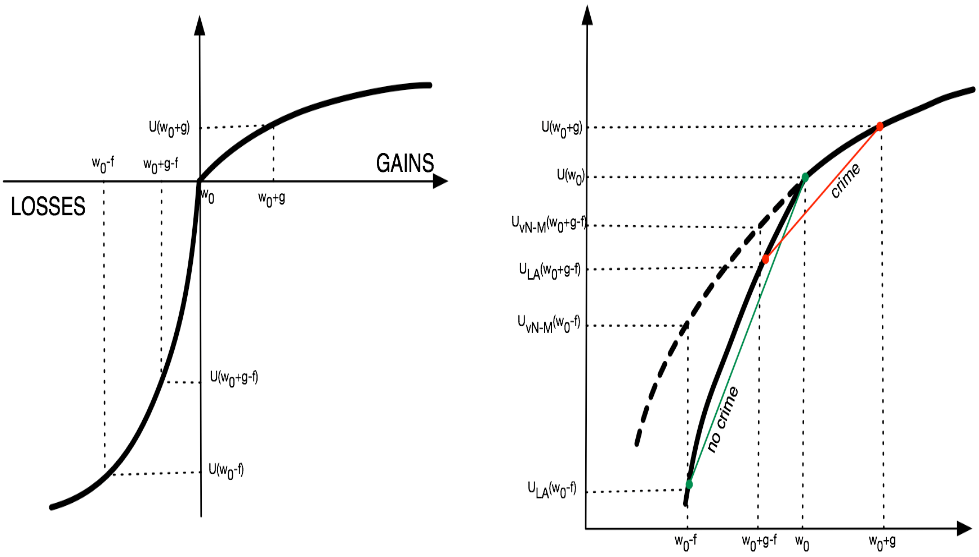
2.2 Loss aversion

The expected utility framework has failed to account for a large body of experimental evidence (Kahneman, 2003; Kahneman et al., 1990). In addition, several phenomena observed in experimental settings, consistent with risk-aversion, cannot be explained by decreasing marginal returns of wealth, but are likely to be due to other factors(Rabin, 2000; Rabin and Thaler, 2001). One of the main violations of the expected utility framework is loss aversion (Kahneman et al., 1991). Loss aversion derives from the fact that people are often framed to think of possible outcomes relative to a certain reference point, rather than as an absolute outcome. This may explain the observed tendency for people to prefer the avoidance of losses (outcomes below the reference point) than the acquisition of comparable gains (outcomes above the reference point). Cumulative prospect theory accounts for loss aversion and also for other behavioral regularities, such as the tendency to overweight extreme, but unlikely events, and the reflection effect (Bowles, 2004).[12] The

[12] Kahneman and Tversky (1979) observed that when decision problems involve not just possible gains, but also possible losses, people's preferences over negative prospects are usually a mirror image of their preferences over positive prospects. While they are risk-averse over prospects involving gains, people become risk-loving over prospects involving losses. This observation is reflected in the convexity of the value function in the losses.

reference-dependent value function typically used in prospect theory is depicted in Figure 3 (note that the reference is w_0 and $f > g$.). It is reasonable to assume that the reference point is centered on the status quo before crime (w_0). The function is kinked at the status quo with loss aversion coefficient of about 2 or more.[13]

Figure 3: Reference-dependent value function and loss aversion



Note: Value function as envisaged by prospect theory (left) and comparison with standard utility function (right).

Loss aversion is a behavioral concept that can be disentangled from both probability weighting and the reflection effect (they are typically combined in the value function used in prospect theory). Indeed, most of the literature on loss aversion employs standard utility to embody loss aversion (Schmidt and Zank, 2005). In the right panel of Figure 3, we compare the utility function with loss aversion and the reference point at w_0 with the utility function described above. Given the concavity of the curve and the kink at w_0 we can observe that

$$\begin{aligned}[U_{vN-M}(w_0) - U_{vN-M}(w_0 - f)] - [U(w_0 + g) - U_{vN-M}(w_0 + g - f)] \\ < [U_{LA}(w_0) - U_{LA}(w_0 - f)] - [U_{LA}(w_0 + g) - U_{LA}(w_0 + g - f)] \quad (5)\end{aligned}$$

Equation (5) indicates that loss aversion provides an additional reason, in addition to risk aversion, for ε_1 to have a larger adverse impact on deterrence than ε_2.

[13] This coefficient is usually measured as $\frac{-U'(-1)}{U'(1)}$. This implies that the utility function immediately to the left of the reference point is at least twice as steep as to the right. For other measures of loss aversion see Köbberling and Wakker (2005).

Proposition 3. *In the presence of loss aversion, type-I errors are more detrimental to deterrence than type-II errors.*

2.3 Type-I error aversion

Economists tend to have an instrumental view of the law: the law is a set of incentives that constrain individuals only as long as it is optimal for them to abide. However, this approach is not commonly shared, and most law scholars and philosophers emphasize the expressive function of rules:[14] the law prescribes a certain behavior and people tend to follow its precepts because it is the "right thing to do", with little regard to the sanction that backs the rule.

In this perspective, ε_1 and ε_2 have very different meanings. When a wrongful acquittal occurs, the violation of the prescribed behavior is not sanctioned, but at the same time the prescription is not questioned. Instead, with a wrongful conviction, a certain behavior is first dictated and then reprimanded, so that the prescription is neglected by the sanction. Therefore, while type-II errors preserve the expressive function of the rule, as its violation is not sanctioned but the precept is unshaken, type-I errors disrupt the expressive function of the rule, as punishing a law-abiding individual necessarily neglects the precept.

We formulate the hypothesis that the disruption of this expressive function caused by type-I errors implies a specific cost for the individual. Individuals are adverse to type-I errors because they impose an additional cost in terms of loss of guidance and motivational crowding-out. As discussed above, it is reasonable to assume that most people, most of the time, look at the law as a set of *guidance rules.* Therefore, the possibility of being convicted assumes a different connotation for the individual who *lets the law guide his behavior.* Beyond the cost of punishment, the law-abiding individual also pays the cost of the loss of guidance, since the law commands to do something and nevertheless punishes him. In addition, wrongful punishment can crowd-out the intrinsic motivation to obey the law.

Let us assume that ε_1-averse agents not only benefit from material payoffs, but also suffer from the occurrence of ε_1 errors, because of the loss of guidance and the feeling of injustice they have been victim of. These agents have a utility function that can be described as $EU_I - \lambda\varepsilon_1$, where λ is a parameter

[14] For economic approaches to the expressive function of the law see Cooter (1998). Indeed, not all economists have a Beckerian view of the law. To begin with, note that many microeconomics textbooks, and generally many works in economics, assume agents to optimize subject to the constrains of the law (see Nance, 1997, note 82, for some examples). On the other hand, the assumption of unconstrained optimization has come under increasing criticisms theoretically (see Harrison, 1986), empirically (e.g. Ellickson, 1991) and behaviorally (see Galbiati and Vertova, 2008, who show how the behavior of agents in the laboratory changes following a change of command without any change in the sanction).

that captures the sensitivity to type-I errors. The utility for the crime choice is instead modelled as before:

$$\begin{cases} EU_I = \varepsilon_1 U(w_0 - f) + (1-\varepsilon_1)U(w_0) - \lambda\varepsilon_1 \\ EU_C = \varepsilon_2 U(w_0 + g) + (1-\varepsilon_2)U(w_0 + g - f) \end{cases} \tag{6}$$

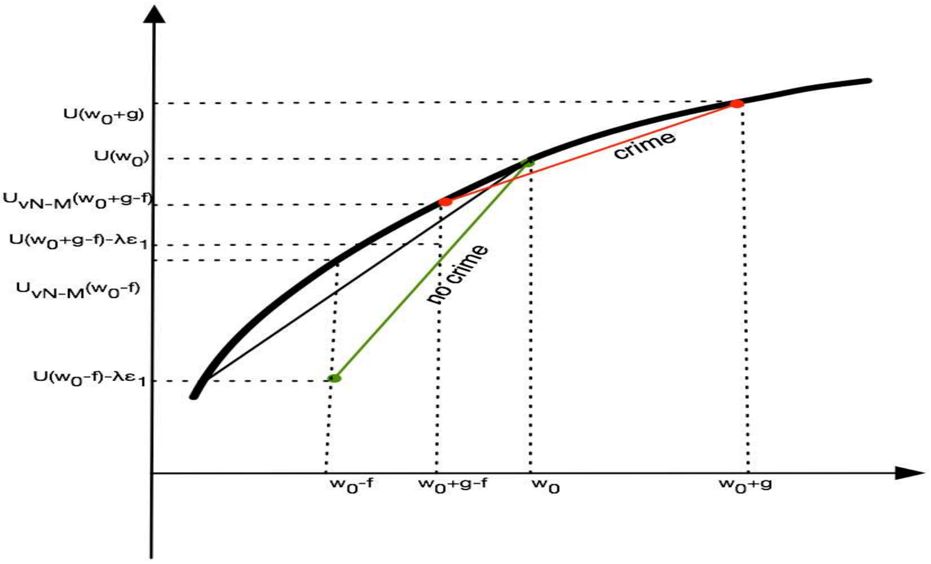
Individuals are therefore deterred if

$$U(w_0) - U(w_0 + g - f) - \varepsilon_1[U(w_0) - U(w_0 - f) + \lambda] \\ - \varepsilon_2[U(w_0 + g) - U(w_0 + g - f)] > 0 \tag{7}$$

As shown in Figure 4, even without assuming concavity of the utility function, type-I-error aversion implies that $[U(w_0)-U(w_0-f)+\lambda] > [U(w_0+g) - U(w_0+g-f)$. As a consequence, ε_1 has a larger adverse impact on deterrence than ε_2.

Proposition 4. *In the presence of type-I-error aversion, type-I errors are more detrimental to deterrence than type-II errors.*

Figure 4: Effects of type-I error aversion



Note that, for those agents that do not commit the crime, the costs of ε_1 in terms of "injustice" or "lost guidance" add up to the costs of ε_1 in terms of "jeopardized deterrence". Conversely, agents opting for the crime do not bear these costs. Therefore, when $f = f^*$, type-I-error aversion tips the balance in favour of the criminal option since it crowds-out the incentives to abide by the law. This effect can be present in addition to the effects of risk-aversion and loss aversion discussed above.

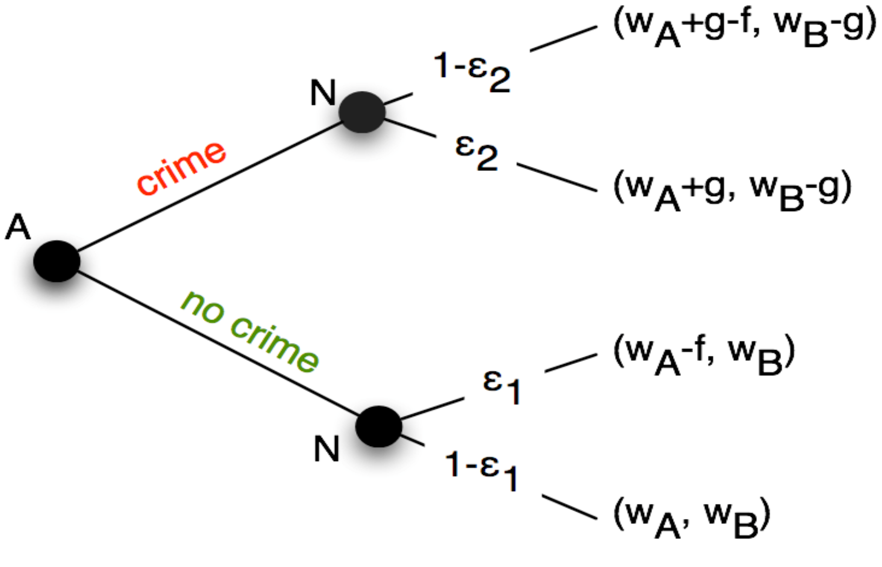
3 Experimental design

This section presents the experimental design. We start by describing the baseline game used to simulate crime in the lab. We then present the different treatments and the hypotheses to be tested. Finally, we describe the experimental procedures.

3.1 Baseline game

The kind of crime mimicked in this experiment is petty larceny. The experimental task is based on a reverse dictator game, described in Figure 5. Two agents, A and B, are randomly matched and assigned an initial endowment (w_A, w_B). Agent A has to decide whether to subtract a sum g from B's endowment. If A decides to take g, there is a probability $1-\varepsilon_2$ that the transfer is detected and, if this happens, A pays a sanction f (while keeping the amount g). If A decides not to take, he will be (wrongfully) sanctioned with probability ε_1. Note that A does not know the size of B's endowment. This allows us to abstract from issues related to distributional fairness or inequality aversion.

Figure 5: Baseline game



This game has been relatively little studied so far in the experimental literature. In our setting, the decision in the reverse dictator game is not about fairness or distribution, as subjects do not know other subjects' endowment. It is rather about the conformity to the social norm of not stealing, or not committing a crime in general. In particular, with this baseline game we can test if and how type-I and type-II errors affect the willingness to abide by the social norm.

3.2 Treatments

Our experimental design is described in Table 1. The endowment of A is set to $w_A = 10$ in treatments T1 to T4. The endowment of B, unknown to subjects A, is set to $w_B = 15$ in all treatments. The amount that can be stolen from B is set to $g = 10$ in all treatments. The key treatment variables are the probabilities of type-I and type-II errors (ε_1, ε_2). They are both set to 0 in the control treatment (T1), that provides a benchmark with optimal deterrence. The probabilities of judicial errors are increased independently in treatments T2 ($\varepsilon_1 = 0$, $\varepsilon_2 = 0.5$) and T3 ($\varepsilon_1 = 0.5$, $\varepsilon_2 = 0.$), respectively, and jointly in treatment T4 ($\varepsilon_1 = 0.25$, $\varepsilon_2 = 0.25$). In treatments T5 and T6 we replicate treatments T2 and T3, while varying the endowment of subject A (w_A), in order to assess the role of risk aversion.

Figure 6 provides a visual comparison of expected utility across the six treatments. Note that, as shown in Table 1, the expected gain from committing the crime is the same ($E\pi_C - E\pi_I = 5$) in treatments T2 to T6. Also note that, relative to the control treatment (T1) the difference in the expected utility obtained from crime and innocence (ΔEU in Table 1) is the same in treatments T3 ($\varepsilon_1 = 0.5$, $\varepsilon_2 = 0.$) and T5 ($\varepsilon_1 = 0$, $\varepsilon_2 = 0.5$). Likewise, the difference in the expected utility obtained from crime and innocence is the same in treatments T2 ($\varepsilon_1 = 0$, $\varepsilon_2 = 0.5$) and T6 ($\varepsilon_1 = 0.5$, $\varepsilon_2 = 0.$).

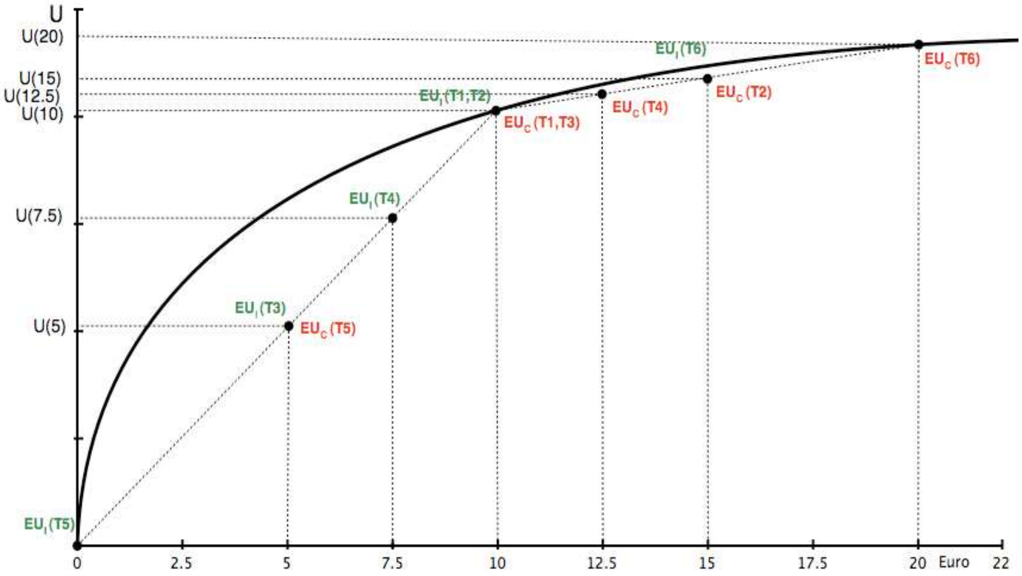
Table 1: Experimental design: comparison of treatments

	T1	T2	T3	T4	T5	T6
ε_1	0	0	0.5	0.25	0	0.5
ε_2	0	0.5	0	0.25	0.5	0
w_A	10	10	10	10	0	20
$E\pi_I$	10	10	5	7.5	0	15
$E\pi_C$	10	15	10	12.5	5	20
EU_I	$U(10)$	$U(10)$	$\frac{U(0)+U(10)}{2}$	$\frac{U(0)+3U(10)}{4}$	$U(0)$	$\frac{U(10)+U(20)}{2}$
EU_C	$U(10)$	$\frac{U(10)+U(20)}{2}$	$U(10)$	$\frac{3U(10)+U(20)}{4}$	$\frac{U(0)+U(10)}{2}$	$U(20)$
ΔEU	0	$\frac{U(20)-U(10)}{2}$	$\frac{U(10)-U(0)}{2}$	$\frac{U(20)-U(0)}{4}$	$\frac{U(10)-U(0)}{2}$	$\frac{U(20)-U(10)}{2}$

Note: ε_1 = probability of type-I error, ε_2 = probability of type-II error, w_A = endowment of subject A, $E\pi_I$ = A's expected payoff if innocent, $E\pi_C$ = A's expected payoff if criminal, EU_I = A's expected utility if innocent, EU_C = A's expected utility if criminal, ΔEU = Net expected utility gain from criminal activity.

This design presents some noteworthy advantages: it has a very simple and intelligible structure; it mimics very intuitively a petty theft; it allows us to disentangle the incentive effect of adjudicative errors from other factors

Figure 6: Comparison of treatments



that may affect subjects' behavior. In addition, the passivity of B removes all strategic uncertainty from A's choice.

We implement the six treatments using a within-subjects design, so that in a session each subject plays the baseline game in 6 different versions, corresponding to the 6 treatments described in Table 1. The experiment is run in four sessions. Within each session, treatments T2,T3,T5 and T6 are played twice, in a different sequence, in order to provide an assessment of the robustness of the key treatment effects. This implies that in each session the reverse dictator game is played in 10 different phases, with a sequence described in Table 2.

Table 2: Sequence of treatments within sessions

Phase	1	2	3	4	5	6	7	8	9	10
Session 1	1	4	2	3	5	6	2	3	5	6
Session 2	4	1	3	2	6	5	3	2	6	5
Session 3	1	4	2	3	5	6	2	3	5	6
Session 4	4	1	3	2	6	5	3	2	6	5

Note: sessions 1-2=Crime Framing with Loss; session 3=Neutral Framing with Loss; session 4=Neutral Framing with no Loss.

In order to assess the role of loss aversion and procedural fairness, we also vary the framing of the game across sessions. In sessions 1 and 2 the game is explained with a language that lets the subjects perceive that the sanction is the punishment for a theft, and indicates w_A as a reference point, so that type-I errors should be perceived as losses (Punishment Framing with Loss).

In session 3, we use a neutral framing for the sanction (Neutral Framing with Loss). In session 4, in addition to neutral framing for the sanction, all payoffs are expressed as potential gains, thus not offering a reference point that leads to potential losses (Neutral Framing with no Loss).

3.3 Hypotheses

Let us define Z_i as the fraction of agents that opt for crime in the population in treatment i. The first question we address is the effect of changes in the probability of type-II errors on deterrence, examining what happens when, ceteris paribus, the probability of detection for criminals is exogenously decreased (Becker's deterrence hypothesis). This can be assessed by comparing $T1$ with $T2$:

> **Hypothesis 1a -** Crime should increase as the expected sanction becomes suboptimal due to a rise in ε_2

$$H_0 : Z_{T2} = Z_{T1} \text{ vs } H_1 : \ Z_{T2} > Z_{T1} \tag{H1a}$$

The second hypothesis we test is based on the Png (1986) extension of the deterrence hypothesis to type-I errors. We examine the effect of type-I errors on deterrence by comparing $T3$ with $T1$. When type-I errors increase, the returns from being innocent decrease, so that Z is expected to rise:

> **Hypothesis 1b -** Crime should increase as the expected sanction becomes suboptimal due to a rise in ε_1

$$H_0 : Z_{T3} = Z_{T1} \text{ vs } H_1 : \ Z_{T3} > Z_{T1} \tag{H1b}$$

Third, we consider the effect of a joint increase in both type-I and type-II errors, by comparing $T1$ with $T4$ (Becker's and Png's deterrence hypothesis). The sanction is suboptimal in $T4$ while it is optimal in $T1$:

> **Hypothesis 1c -** Crime should increase as the expected sanction becomes suboptimal due to a rise in both ε_1 and ε_2
>
> $$H_0 : Z_{T1} = Z_{T4} \text{ vs } H_1 : \ Z_{T4} > Z_{T1} \tag{H1c}$$

The next question we address is whether type-I and type-II judicial errors have the same impact on deterrence. The standard theory of crime deterrence, under risk neutrality, predicts that both errors are equally detrimental to deterrence. We can test this hypothesis by comparing $T2$ with $T3$. These two treatments differ only with respect to the probability of the two types of judicial errors, while the expected gains from crime are the same in both treatments:

> **Hypothesis 2 -** Crime should increase in the same way in response to a given increase in ε_1 and ε_2:

$$H_0 : Z_{T2} = Z_{T3} \text{ vs } H_1 : \; Z_{T2} < Z_{T3} \tag{H2}$$

When comparing treatments T2 and T3, a difference in the proportion of criminals could be explained by a number of factors. A first explanation could be linked to risk aversion. Although the expected net gains from crime in the two treatments is the same, the wealth of the individual is not. This implies that type-II errors could have a smaller effect on deterrence than type-I errors just because of diminishing marginal returns if the utility function is concave. Second, if subjects are loss averse, the effect of judicial errors on deterrence can be asymmetric. In the case of type-I errors, the riskiness of the uncertain outcome increases the *loss* of expected utility for an innocent. In the case of type-II errors, the riskiness of the uncertain outcome reduces the *gain* in expected utility for a criminal.[15] As a consequence, for a given expected sanction, in the presence of loss-aversion type-I errors should have a larger negative effect on deterrence than type-II errors. Third, the two type of errors could be perceived differently in terms of procedural fairness. If type-I errors are perceived as less fair than type-II errors, this leads to a higher incidence of crime despite an equal expected sanction.

We test for the effect of risk attitudes by comparing $T2$ with $T5$ and $T3$ with $T6$. In the presence of risk aversion, we expect the effect of a 0.5 increase in ε_2 to be stronger in T5, where $w_A = 0$, than in T2, where $w_A = 10$. Similarly, risk aversion implies that the effect of a 0.5 increase in ε_1 is stronger in T3, where $w_A = 10$, than in T6, where $w_A = 20$. These exogenous changes in w_A allow us to compare the effects of type-I and type-II errors while controlling for risk attitudes. We thus compare $T2$ with $T6$ and $T3$ with $T5$. Note that there is the same difference in expected utility between honesty and crime within each pair of treatments (see Figure 6). Therefore, in both cases, the effect of ε_1 and ε_2 on deterrence should be the same, irrespective of attitudes towards risk, as the difference in expected utility is the same in both treatments.

> **Hypothesis 3 -** Controlling for the effect of attitudes towards risk, crime should increase in the same way in response to a given increase in ε_1 or ε_2.

We thus test whether type-I and type-II errors have the same effect on deterrence, while controlling for risk attitudes, setting our null hypotheses as follows:

[15] Of course if $f > g$ there could potentially be some losses in case of crime as well. However, while with $\varepsilon_1 > 0$ the innocent surely bears an expected loss, with $\varepsilon_2 > 0$ the guilty might have an expected gain.

$$H_0 : Z_{T2} = Z_{T6} \text{ vs } H_1 : \; Z_{T2} < Z_{T6} \tag{H3a}$$

$$H_0 : Z_{T3} = Z_{T5} \text{ vs } H_1 : \; Z_{T3} > Z_{T5} \tag{H3b}$$

A stronger adverse effect of type-I errors on deterrence could be reflecting either loss aversion or type-I error aversion. In order to disentangle the two effects, we compare the behavior of subjects in different sessions, exploiting the differences in the framing of the experimental task. Framing A, used in sessions 1 and 2, describes punishment with a pronounced criminal frame. It also offers the initial endowment as a reference point and the two choices produce either a loss (honesty) or a gain (crime). In this case, both loss aversion and type-I error aversion may play a role. Framing B, used in session 3, describes the sanction more neutrally, but still suggests a reference point that shapes the two choices in terms of loss or gain. With this framing, since we have neutralized type-I aversion, we should observe less pronounced differences between Z_{T2} and Z_{T6}, and between Z_{T3} and Z_{T5}, as compared to framing A. We thus test whether ε_1 and ε_2 have the same effect on deterrence, while controlling for risk attitude and for type-I error aversion by using a neutral framing. Framing C, used in session 4, describes the choice without the criminal frame and without the loss frame. Controlling for the effect of risk aversion, type-I error aversion, and loss aversion, we test whether crime increases in the same way in response to a given increase in ε_1 or ε_2.

3.4 Procedures

The experiment was implemented in four sessions, with 24 subjects participating in each session, for a total of 96 subjects. In each session, subjects were randomly assigned to a computer terminal at their arrival. In order to ensure public knowledge, instructions were distributed and read aloud (see the Appendix). Sample questions were distributed to ensure understanding of the experimental procedures. Answers were privately checked and, if necessary, explained to the subjects, and the experiment did not start until all subjects had answered all questions correctly.

Within each session, subjects were matched in pairs in each phase, as described in Table 2, using a perfect stranger matching mechanism. Participants were informed that each subject would play the role of player A, but at the end of the experiment roles A and B would be randomly determined within pairs, so that ex post each subject only played one role. The experiment employed a no-feedback design, so that we could elicit the relevant responses from all subjects and obtain statistically independent observations across subjects. Subjects were paid on the basis of the outcome of one randomly picked phase and one randomly selected role out of the two they had played.

The experiment was conducted in the Experimental Economics Laboratory of the University of Milan Bicocca in April 2009. Participants were undergraduate students of Economics recruited by e-mail using a list of voluntary potential candidates. None of the subjects had participated previously in similar reverse dictator games. No show-up fee was paid. Payments ranged between 0 and 20 euro, and the average payment was 11.1 euro for sessions lasting on average approximately 45 minutes. The experiment was run using the experimental software z-Tree (Fischbacher, 2007).

4 Results

Figure 7 displays the percentage of criminals, by phase, in sessions 1 and 2. In T1, where there is optimal deterrence in the absence of judicial errors ($\varepsilon_1 = \varepsilon_2 = 0$), the percentage of criminals is 29 per cent. This figure provides the benchmark against which to assess the effects of judicial errors on deterrence. It is interesting to observe that the percentage of criminals is about 29 per cent in both sessions, despite the sequence of treatments being reversed. Similar results, irrespective of the sequence of treatments, are also obtained for treatment T4 ($\varepsilon_1 = \varepsilon_2 = 0.25$): 75 and 79 per cent in sessions 1 and 2, respectively. This indicates that the order of treatments does not play a major role.

Figure 7: Percentage of criminals by phase, sessions 1 and 2

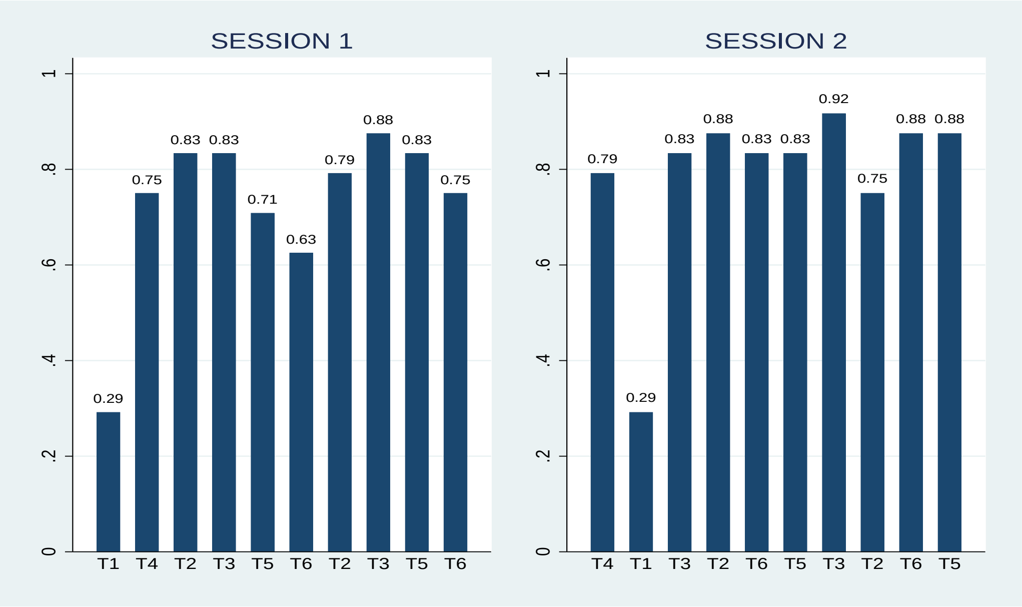


Table 3 displays the percentage of criminals in sessions 1 and 2, by treatment, both overall and by subset of phases (3-6 and 7-10). The figures indicate that judicial errors have a large negative impact on deterrence. When both ε_1 and ε_2 are increased jointly to 0.25, the percentage of criminals rises to 77.1 per cent. This effect is strongly statistically significant, on the basis of a two-sided null hypothesis and 48 independent paired observations (McNemar's $\chi^2_1 = 17.6$, p-value = 0.00). The effect is indeed stronger when, keeping constant the expected gains from crime, judicial errors are increased individually. In T2, where ε_2 is increased to 0.5, the percentage of criminals rises to 81.2 per cent overall. In T3, where ε_1 is increased to 0.5, the average percentage of criminals rises to 86.4 per cent overall. In both cases, focusing on either phases 3-6 or 7-10, the effect is strongly statistically significant. These results indicate that judicial errors have a large and significant effect on crime deterrence. They also indicate that, for a given expected gain from crime, the effect is stronger when judicial errors are increased individually rather than jointly.

Table 3: Percentage of criminals by treatment, sessions 1 and 2

	T1	T4	T2	T3	T5	T6
Overall	29.2	77.1	81.2	86.4	81.2	77.1
Phases 3-6			85.4	83.3	77.1	72.9
Phases 7-10			77.1	89.6	85.4	81.2

Note: see Table 1 for a description of parameter calibration within treatments.

> **Result 1 -** A higher probability of either type-I or type-II error causes a large and significant increase in crime.

We now turn to hypothesis 2: Do type-I and type-II judicial errors have the same impact on deterrence? We test this hypothesis by comparing the average percentage of criminals in $T2$ and $T3$. Averaging across all phases, there is a positive difference of 5.2 percentage points between crime rates in T3 and T2. Focusing on the last four phases, the difference is much larger, as the percentage of criminals is 77.1 and 89.6 per cent in T2 and T3, respectively. For the relevant one-sided hypothesis, on the basis of 48 independent paired observations, the difference is strongly statistically significant (McNemar's $\chi^2_1 = 3.6$, p-value = 0.029). Contrary to the predictions of the standard theory of crime deterrence, type-I and type-II errors do not have symmetric effects on crime deterrence.

> **Result 2 -** The probability of type-I errors has a larger impact on deterrence than the probability of type-II errors.

In order to assess the role played by risk attitudes in explaining the differences in the effects of type-I and type-II judicial errors on deterrence, we vary the subjects' endowments so as to obtain the same difference in expected utility between honesty and crime in treatments $T2$ and $T6$ and $T3$ and $T5$, respectively. It is interesting to observe that the overall percentage of criminals falls from 86.4 in T3 ($w_A = 10$) to 77.1 in T6 ($w_A = 20$). This effect is statistically significant at the 5 per cent significance level both in phases 3-6 and 7-10. The percentage of criminals is instead unchanged at 81.2 in T2 ($w_A = 10$) and T5 ($w_A = 0$) overall, but it rises from 77.1 in T2 to 85.4 in T5 when focusing on the last four phases, although this effect is only marginally significant (p-value = 0.10 for a two-sided hypothesis). Overall, these results indicates that risk aversion plays a significant role for subjects' decisions.

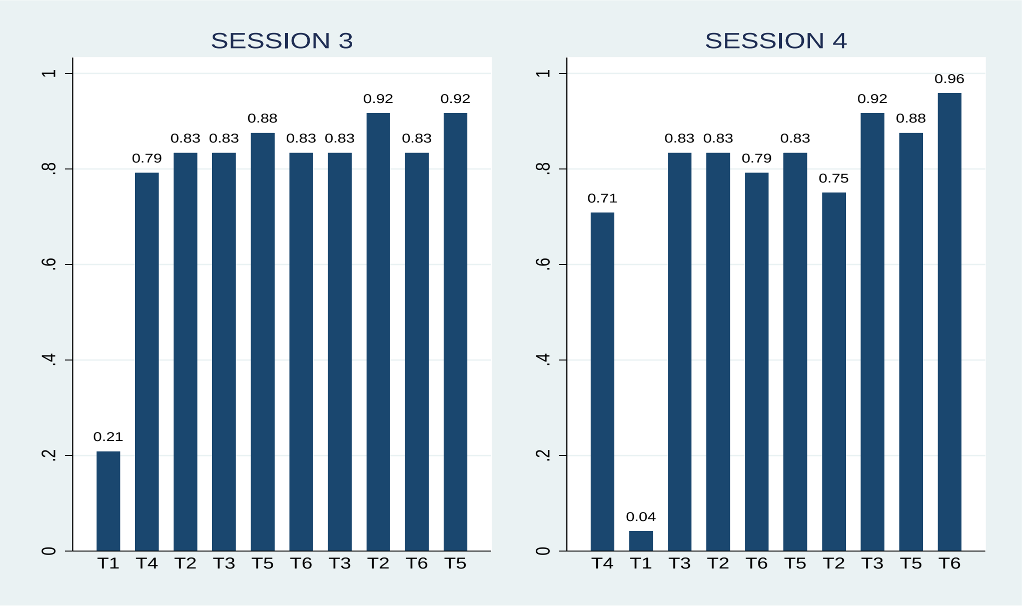
We therefore turn to comparing the effects of type-I and type-II judicial errors, while keeping constant the difference in expected utility between crime and honesty, so as to control for the effects of risk attitudes. Overall, the percentage of criminals is 77.1 in T6 ($\varepsilon_1 = 0.5$) and 81.2 in T2 ($\varepsilon_2 = 0.5$), while 86.4 in T3 ($\varepsilon_1 = 0.5$) and 81.2 in T5 ($\varepsilon_2 = 0.5$). Focusing on the last four phases, the difference between T6 (81.2) and T2 (77.1) is positive, but relatively small and not statistically significant (McNemar's $\chi^2_1 = 0.33$, p-value = 0.56). Similarly, the difference in the percentage of criminals between T3 and T5 is positive (4.2 percentage points) but small and not statistically significant (McNemar's $\chi^2_1 = 1.00$, p-value = 0.31). Overall, these results indicate that, once we control for risk attitudes, the effect of type-I errors on deterrence is only marginally stronger than that of type-II errors, and the difference is no longer statistically significant.

> **Result 3 -** Controlling for risk attitudes, the effect of type-I errors on deterrence is marginally stronger than the effect of type-II errors, but the difference is not statistically significant.

Finally, we turn to assessing the role played by loss aversion and procedural fairness in explaining the asymmetric effects of judicial errors on deterrence. Figure 8 displays the percentage of criminals, by phase, in sessions 3 and 4. Table 4 displays the corresponding figures by treatment, both overall and by subset of phases. The overall percentage of criminals in T2 rises from 81.1 in sessions 1-2 to 87.5 in session 3. On the contrary, in T3 it falls from 86.4 in sessions 1-2 to 83.3 in session 3. As expected, whereas in the presence of type-II errors the framing of the sanction as a punishment for a theft has a positive effect on crime deterrence, in the presence of type-I errors the effect of framing on deterrence is small and negative. These results, albeit at a qualitative level, are consistent with the hypothesis of type-I error aversion.

Focusing on the results for session 3, the larger effect of ε_1 on deterrence, relative to ε_2, virtually disappears when a neutral framing is used. The sign of the difference between the effects of type-I and type-II errors, keeping constant the difference in expected utility between crime and honesty, is indeed reversed. The overall percentage of criminals is 87.5 and 83.3 in T2 and T6, respectively, and 83.3 and 89.6 in T3 and T5, and these differences are not statistically significant. Focusing on the results for session 4, the

Figure 8: Percentage of criminals by phase, sessions 3 and 4



overall percentage of criminals rises to 87.5, relative to 83.3 in session 3. This positive effect, although not statistically significant, is not consistent with the theoretical predictions on the role of loss aversion for the effects of judicial errors.

Overall, although only at a qualitative level, the results for sessions 3 and 4 indicate that type-I error aversion – a differential sensitivity to procedural fairness for the two types of judicial errors – may be playing a role in explaining the asymmetric effects of ε_1 and ε_2 on crime deterrence. On the contrary, loss aversion does not seem to have a significant effect.

Table 4: Percentage of criminals by treatment, sessions 3 and 4

	T1	T4	T2	T3	T5	T6
Unframed (s3)						
Overall	21.8	79.1	87.5	83.3	89.6	83.3
Phases 3-6			83.3	83.3	87.5	83.3
Phases 7-10			91.7	83.3	91.7	83.3
Unframed, no loss (s4)						
Overall	4.1	70.8	79.1	87.5	85.4	87.5
Phases 3-6			83.3	83.3	83.3	79.1
Phases 7-10			75.0	91.7	87.5	95.8

Note: see Table 1 for a description of parameter calibration within treatments.

5 Conclusions

Judicial errors against innocent defendants represent an important issue not only theoretically, but also empirically. Although no country produces official statistics on judicial errors, there is some evidence produced by various national and international organizations.[16] Gross et al. (2005) emphasize how most relevant data sets are based on exoneration cases:[17] their own data base lists 340 exonerations in the US between 1989 and 2003. But counting only the people that eventually obtain an official exoneration – they argue – may be only a glimpse of the "total number of miscarriages of justice in America [that] in the last fifteen years must be in the thousands, perhaps tens of thousands."

The first aim of this paper was to extend the economic theory of crime deterrence to better account for the role type-I errors. In particular, the paper cast light on the asymmetric effects of the two types of judicial errors: wrongful convictions were shown to be more detrimental to deterrence than wrongful acquittals. We argued that both risk-aversion and loss-aversion may contribute to produce this asymmetry. Moreover, we formulated the hypothesis that an individual might be particularly averse to type-I errors, because of the specific effect that wrongful convictions have on the expressive function of the law.

The second aim of the paper was to provide an empirical test of the predictions of our extended model using a laboratory experiment. The experimental design was based on a reverse dictator game, simulating a simple crime such as petty larceny. We exogenously manipulated the probability of type-I and type-II errors, and compared across treatments the propensity of individuals to commit the crime. The findings are largely consistent with our theoretical predictions. Both types of judicial errors have a large and significant impact on deterrence. However, wrongful convictions are found to have a stronger negative impact on deterrence than wrongful acquittals. The analysis indicates that this result is largely explained by risk aversion. We also find some evidence consistent with the hypothesis that type-I error aversion may be playing a role in explaining the asymmetric effects of type-I and type-II judicial errors on deterrence.

It must be emphasized that, although in this paper the argument was mainly framed in the context of criminal law, judicial errors can occur in any adjudicative process. Our analysis can therefore be relevant for the

[16] *Forejustice*, for instance, provides a list of 2803 cases of innocents convicted (see also Fon and Schaefer, 2007). This list envisages cases from 84 countries. In these 2803 cases, 184 people were executed, 514 people were sentenced to death, 575 people were sentenced to life in prison. 2,622 people on this list were judicially exonerated or pardoned and 414 people were exonerated after a false confession (187 in the US). (http://www.forejustice.org, accessed on July, 16th 2009).

[17] Cases for which people have been first convicted (and the conviction reached the final stage often including the court of last resort), and then, after some time, declared innocent because some new evidence has emerged.

decisions taken by organizations as diverse as administrative agencies, commercial trade associations, religious bodies, professional sports leagues and the like.[18]

Overall, the behavioral implications of the deterrence hypothesis pose important challenges to the economic theory of the public enforcement of law. Type-I errors may jeopardize deterrence more than it has been so far predicted in the literature. This makes a strong economic case for the public authority to place particular emphasis on type-I errors, since the economic goal of public law enforcement is to achieve optimal deterrence.

The results presented in this paper also have important implications for the law and economics literature. In particular, they help the standard model of optimal crime deterrence to make sense of the pro-defendant bias observed in the criminal procedures of modern democracies. More generally, they contribute to explain the common wisdom that the conviction of an innocent should be considered far worse than the acquittal of a guilty individual.

[18] Consider, for instance, the case of tax audit, an adjudicative process that is often prone to type-I errors, not least because tax agencies pursue conflicting interests. Tax audit errors are notoriously numerous in some countries.

Appendix - Instructions

Welcome and thanks for participating in this experiment. During the experiment you are not allowed to talk or communicate in any way with other participants. If at any time you have any questions raise your hand and one of the assistants will come to you to answer it. By following the instructions carefully you can earn an amount of money that will depend on your choices and the choices of other participants. At the end of the experiment, the resulting amount will be paid to you in cash.

General rules

- There are 24 subjects participating in this experiment.
- The experiment takes place in 10 independent phases. Instructions for each phase will appear on the screen.
- In each phase, 12 couples of two participants will be formed randomly and anonymously, so that in each phase you will interact with a different subject.
- Within each couple, the two subjects will be randomly assigned two different roles: A and B.
- Therefore, in each phase each subject will interact exclusively with the other subject in her pair, without knowing her/her identity, with the role (A or B) assigned with equal probability.
- The choices that you and the other subject will make in each and the corresponding outcomes will be communicated at the end of the experiment.
- At the end of the experiment only one of the 10 phases will be selected randomly and earnings for each participant will be determined on the basis of the selected phase.

How earnings are determined within each phase

- Your choice consists in deciding whether to steal 10 euro to the subject you have been paired with.
- You will have a known endowment of x euro, whereas the other subject will have an endowment of y euro (unknown to you).
- After you have made your choice, there will be a check in order to punish those who steal.
- The check will be subject to a certain error, as follows:

  - If you have stolen 10 euro to the other subject, you will have to pay a 10 euro fine with a probability of p.
  - If you have not stolen 10 euro to the other subject, you will have to pay a 10 euro fine with a probability of q.

- Therefore, within each phase, earnings will be determined as follows:
  - If you decide to steal, the other subject will earn y −10 euro and you will earn:
    x euro with probability p
    $x + 10$ euro with probability $1 - p$
  - If you decide not to steal, the other subject will earn y euro and you will earn:
    $x - 10$ euro with probability q
    x euro with probability $1 - q$

How final earnings are determined

- At the end of the experiment only one of the 10 phases will be selected randomly and the earnings for each participant will be determined on the basis of the selected phase.
- Within each phase each subject will have both an active role and a passive role. At the end of the experiment the earnings of each subject will be determined on the basis of one of the two roles, to be selected randomly.

[End of instructions]

Framing across sessions

In order to illustrate the differences in the formulation of instructions among the sessions, we report below the description of T4 under each framing:

Sessions 1 and 2 (Crime Framing with Loss). In this phase you have an initial endowment of 10 euro. If you **steal** 10 euro from the other subject, you have to pay a fine of **10 euro** with probability 75%. Therefore, if you steal, you will gain all in all i) **10 euro** with probability 75% and ii) **20 euro** with probability 20%.

If you **do not steal** 10 euro from the other subject, you have to pay a fine of **10 euro** with probability 20%. Therefore if you do not steal, you will gain all in all i)**5 euro** with probability 20% and ii)**10 euro** with probability 75%.

Choice: **do you want to steal 10 euro from the other subject?**

Session 3 (Neutral Framing with Loss). In this phase you have an initial endowment of 10 euro. If you **steal** 10 euro from the other subject, you loose **10 euro** from the bet with probability 75%. Therefore, if you steal, you will gain all in all i) **10 euro** with probability 75% and ii) **20 euro** with probability 20%.

If you **do not steal** 10 euro from the other subject, you loose **10 euro** from the bet with probability 20%. Therefore if you do not steal, you will gain all in all i)**5 euro** with probability 20% and ii)**10 euro** with probability 75%.

Choice: **do you want to steal 10 euro from the other subject?**

Session 4 (Neutral Framing with no Loss) . In this phase you have an initial endowment of 0 euro. If you **steal** 10 euro from the other subject, you have to pay a fine of **10 euro** with probability 75%. Therefore, if you steal, you will gain all in all i) **10 euro** with probability 75% and ii) **20 euro** with probability 20%.

If you **do not steal** 10 euro from the other subject, you have to pay a fine of **10 euro** with probability 20%. Therefore if you do not steal, you will gain all in all i)**5 euro** with probability 20% and ii)**10 euro** with probability 75%.

Choice: **do you want to steal 10 euro from the other subject?**

References

Abbink, K., “Laboratory experiments on corruption,” *The Handbook of Corruption*, 2006.

Backes-Gellner, U., D. Bessey, K. Pull, and S.N. Tuor, “What Behavioural economics teaches personnel economics,” *Institute for Strategy and Business Economics University of Zurich*, 2008.

Becker, Gary, “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, 1968, *76*, 169–217.

Block, Michael K. and Joseph Gregory Sidak, “The Cost of Antitrust Deterrence: Why not Hang a Price Fixer Now and Then?,” *Georgetown Law Journal*, 1980, *68* (5), 1131–1139.

Bowles, Samuel, *Microeconomics: behavior, institutions, and evolution*, Princeton: Princeton University Press, 2004.

Cooter, Robert, “Expressive Law and Economics,” *The Journal of Legal Studies*, 1998, *27* (2, Social Norms, Social Meaning, and the Economic Analysis of Law), 585–608.

Craswell, R. and J.E. Calfee, “Deterrence and uncertain legal standards,” *Journal of Law, Economics, and Organization*, 1986, *2* (2), 279–303.

Demougin, Dominique and Claude Fluet, “Deterrence versus Judicial Error: A Comparative View of Standards of Proof,” *Journal of Institutional and Theoretical Economics*, 2005, *161* (2), 193–206.

Ehrlich, Isaac, “The Optimum Enforcement of Laws and the Concept of Justice: A Positive Analysis,” *International Review of Law and Economics*, 1982, *2*, 3–27.

Ellickson, R. C., *Order Without Law: How Neighbors Settle Disputes*, Harvard University Press, 1991.

Falk, A and U Fischbacher, “"Crime" in the lab-detecting social interaction,” *European Economic Review*, 2002, *46*, 859–869.

Falk, Armin and Simon Gächter, “experimental Labour Economics,” in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, Basingstoke: Palgrave Macmillan, 2008.

Fehr, E. and S. Gachter, “Cooperation and punishment in public goods experiments,” *American Economic Review*, 2000, *90* (4), 980–994.

Feinberg, William E., "Teaching the Type I and Type II Errors: The Judicial Process," *The American Statistician*, 1971, *25* (3).

Fon, V. and H.B. Schaefer, "State Liability for Wrongful Conviction: Incentive Effects on Crime Levels," *Journal of Institutional and Theoretical Economics*, 2007, *163* (2), 269–284.

Galbiati, Roberto and Pietro Vertova, "Obligations and cooperative behaviour in public good games," *Games and Economic Behavior*, 2008, *doi:10.1016/j.geb.2007.09.004.*

Garoupa, Nuno, "The Theory of Optimal Law Enforcement," *Journal of Economic Surveys*, 1997, *11* (3), 267–295.

_ **and Matteo Rizzolli**, "Wrongful Convictions Do Lower Deterrence," *mimeo*, 2009.

Gross, Samuel R., Kristen Jacoby, Daniel J. Matheson, Nicholas Montgomery, and Sujata Patil, "Exonerations in the United States 1989 through 2003," *The Journal of Criminal Law and Criminology*, 2005, *95* (2), 523–560.

Harris, J.R., "On the Economics of Law and Order," *Journal of Political Economy*, 1970, *78* (1), 165.

Harrison, J., "Egoism, Altruism, and Market Illusions: The Limits of Law and Economics," *UCLA Law Review*, 1986, *33* (5), 1309–1354.

Hoerisch, Hannah and Christina Strassmair, "An experimental test of the deterrence hypothesis," *University of Munich, Department of Economics Discussion papers Series*, 2008, *04-08.*

Immordino, Giovanni and Michele Polo, "Judicial Errors and Innovative Activity," *IGIER (Innocenzo Gasparini Institute for Economic Research) Working paper*, 2008, (337).

Kahneman, D. and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–292.

Kahneman, Daniel, "Maps of Bounded Rationality: Psychology for Behavioral Economics," *The American Economic Review*, 2003, *93* (5), 1449–1475.

_ **, Jack L. Knetsch, and Richard H. Thaler**, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 1990, *98* (6), 1325–1348.

_ **,** _ **, and** _ , "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies," *Journal of Economic Perspectives*, 1991, *5* (1), 193–206.

Kaplow, L., "Accuracy in Adjudication," in "The New Palgrave Dictionary of Economics and the Law'," Harvard Law School, 1996.

Kaplow, Louis, "The Value of Accuracy in Adjudication: An Economic Analysis," *Journal of Legal Studies*, 1994, *23* (1), 307–401.

Kaplow, Luis and Steven Shavell, "Accuracy in the Determination of Liability," *Journal of Law and Economics*, 1994, *37* (1), 1–15.

Köbberling, V. and P.P. Wakker, "An index of loss aversion," *Journal of Economic Theory*, 2005, *122* (1), 119–131.

Lando, Henrik, "Does Wrongful Conviction Lower Deterrence?," *Journal of Legal Studies*, 2006, *35* (2), 327–338.

Nance, D. A., "Guidance Rules and Enforcement Rules: A Better View of the Cathedral," *Virginia Law Review*, 1997, *83* (5), 837–937.

Png, Ivan P. L., "Optimal Subsidies and Damages in the Presence of Judicial Error," *International Review of Law and Economics*, 1986, *6* (1), 101–05.

Polinsky, A. Mitchell and Steven Shavell, "The Theory of Public Enforcement of Law," in A. Mitchell Polinsky and Steven Shavell, eds., *Handbook of Law and economics*, Amsterdam: Elsevier, 2007.

Polinsky, A.M. and S. Shavell, *Public Enforcement of Law*, 2nd ed., Vol. The New Palgrave Dictionary of Economics, Palgrave MacMillan, 2008.

Rabin, M and RH Thaler, "Risk Aversion," *Journal of Economic Perspectives*, 2001.

Rabin, Matthew, "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 2000, pp. 1281–1292.

Schmidt, U and H Zank, "What is Loss Aversion?," *Journal of Risk and Uncertainty*, 2005.

Schulze, G.G. and B. Frank, "Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility," *Economics of Governance*, 2003, *4* (2), 143–160.

Sonnemans, Joep and Frans van Dijk, "Errors in Judicial Decisions," Tinbergen Institute Discussion Papers 08-089/1, Tinbergen Institute September 2008.

Strandburg, Katherine, "Deterrence and the Conviction of Innocents," *Connecticut Law Review*, 2003, *35*, 1321–1350.

Torgler, Benno, "Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments," *Journal of Economic Surveys*, 2002, *16* (5), 657–683.

Visser, M.S., W.T. Harbaugh, and N.H. Mocan, "An Experimental Test of Criminal Behavior Among Juveniles and Young Adults," *National Bureau of Economic Research Working paper 12507*, 2006.