Judicial Errors and Crime Deterrence: Theory and Experimental Evidence

Matteo Rizzolli *Free University of Bozen–Bolzano*
Luca Stanca *University of Milan–Bicocca*

Abstract

The economic theory of crime deterrence predicts that the conviction of an innocent individual (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 type I error aversion, type I errors have a stronger effect on deterrence than type II errors. We test these predictions with two experimental studies in which participants choose whether to steal from other individuals, under alternative combinations of probabilities of judicial errors. The results indicate that both types of errors have a significant impact on deterrence. As predicted, type I errors have a stronger impact on deterrence than type II errors. This asymmetry is entirely explained by differences in the expected utility gains from crime, whereas nonexpected utility factors do not play a significant role.

Un coupable puni est un exemple pour la canaille; un innocent condamné est l'affaire de tous les honnêtes gens. [Jean de La Bruyère 1696]

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. [Innocence Project]

1. Introduction

It is common wisdom to consider the punishment of an innocent individual to be more questionable than the acquittal of a guilty individual. As *honnêtes gens,*

We thank Nuno Garoupa, Kristoffel Grechenig, John Hey, Henrik Lando, Murat Mungan, Antonio Nicita, Piero Pasotti, Mitch Polinsky, Alan Schwartz, Henry Smith, Paolo Vanin, and Daniel Zizzo for useful suggestions and discussions. We also thank participants at the 2009 conference of the European Association of Law and Economics, Rome (September 17–19), the 2009 European meeting of the Economic Science Association, Innsbruck (August 17–20), and the 2010 meeting of the

[*Journal of Law and Economics,* vol. 55 (May 2012)]
 0022-2186/2012/5502-0011$10.00

in the words of de La Bruyère (1696), we share this opinion.[1] 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 individual (type I error) is equivalent to 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 2008). The implication of this result is important: if optimal deterrence is what matters, the policy maker—and the judge—should be indifferent between type I and type II errors.

In this paper, we build on the theory of optimal deterrence to show that several factors may explain the 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, as individuals value the avoidance of type I error losses more than the acquisition of equivalent type II error gains. Third, wrongful punishments imply a specific cost in terms of loss of guidance that may result in what we refer to as type I error aversion. All these factors imply that, for a given expected gain from crime, type I errors should have a stronger impact on deterrence than type II errors.

We evaluate these theoretical predictions by means of two laboratory experiments. The objective of the analysis is to test the deterrence hypothesis, focusing on the role of both types of judicial errors, and to assess whether type I and type II errors have different effects on deterrence. Our findings indicate that both type I and type II errors have a large and significant impact on deterrence. Consistent with our hypothesis, type I errors have a stronger effect than type II errors. However, when the expected utility gains from crime are kept constant across treatments, we find no difference between the effects of type I and type II errors. This indicates that the asymmetry in the effects of judicial errors can be explained within the expected utility framework. Nonexpected utility factors, such as loss aversion and type I error aversion, are not found to play a significant role.

The paper is structured as follows. Section 2 briefly reviews the related literature on judicial errors and crime deterrence. Section 3 provides the theoretical framework. Section 4 describes the experimental design. Section 5 presents the

---

American Law and Economics Association, Princeton, N.J. (May 7–8). We thank the Free University of Bozen-Bolzano for financial support.

[1] The translation of de La Bruyère's quote is as follows: "A guilty man punished is an example for the rabble; an innocent man condemned is a matter for all honest people." The Innocence Project is "a nonprofit 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" (http://www.innocenceproject.org). Through the years, the project has managed to obtain the exonerations of 225 people who had received a final sentence, 17 of whom were on death row.

results. Section 6 discusses the interpretation and implications of the findings, and Section 7 concludes.

2. Related Literature

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) models explicitly the effects 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 expected payoff of engaging in the unlawful activity. Thereafter, this extension has been generally incorporated in the economic literature on crime deterrence (see, for example, 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 expected returns from honesty and crime.

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 individual may reinforce deterrence if it is perceived by other imperfectly informed would-be offenders as the conviction of a guilty individual. They may incorrectly interpret a higher frequency of type I errors as a lower frequency 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, 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 error probabilities, may result in either higher or lower levels of deterrence (Kaplow 1998).[2] 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; Dacey and Gallant 1997).[3] More recently, the Png (1986) equivalence result has been questioned by Lando (2006), who argues that when mistakes about the identity of the criminal are taken into account, type I errors have no detrimental effect on deterrence.[4] 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, the deterrence hypothesis has been widely studied (Levitt

[2] The nuisance effect of errors on the perception of the rule in place is pointed out by Craswell and Calfee (1986).

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

[4] 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. On this point, see also the discussion in Garoupa and Rizzolli (2012).

and Miles 2007). A large body of literature has focused on whether policies that affect the magnitude of the expected sanction have an impact on the crime rate. Numbers of arrests and convictions have generally been used as proxies for the expected sanction. These figures, however, do not distinguish between correct and wrongful convictions. More generally, it is difficult to assess the number of wrongful convictions through observational data, since any increase in the number of total convictions can be driven by an increase in crime, an increase in wrongful convictions, or a decrease in wrongful acquittals. As a consequence, relatively little is known empirically about the impact of wrongful convictions on the decision to commit a crime (Smith 2004; Gross and O'Brien 2008; Gould and Leo 2010). The experimental method offers a solution to this problem, as it allows the researcher to exogenously manipulate the probabilities of type I and type II errors.

A small number of experimental studies have recently investigated the deterrence hypothesis in the lab. DeAngelo and Charness (2009) use a game that mimicks traffic law violations. In their setting, however, crime creates no externality and therefore does not pose a social dilemma. Harbaugh, Mocan, and Visser (2011) explore the deterrence hypothesis experimentally using a theft game.[5] 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 a type II error. By varying both the amount to be potentially stolen and the probability of detection, the authors study whether the decision to steal satisfies the generalized axiom of revealed preference. Schildberg-Hörisch and Strassmair (2012) conduct an experiment on crime deterrence using a similar theft game. They examine whether observed decisions to steal are compatible with alternative theories of social preferences and find that the deterrence hypothesis poorly describes observed behavior. When there is a small probability of detection, subjects steal less than they do when the expected sanction is high. However, when the sanction becomes even more severe, the deterrence hypothesis holds. Schildberg-Hörisch and Strassmair conclude that the threat of a sanction may crowd out law-abiding behavior.[6] Rauhut (2009) tests in the lab game-theoretic predictions related to crime deterrence in the presence of strategic interactions between the authority and the potential criminals. Engel and Irlenbusch (2010) and Enukashvili (2010) try to disentangle the general deterrence and special prevention purposes of punishment. Grechenig, Nicklisch, and Thöni (2010) investigate the role of uncertainty (and thus the

[5] Other recent experimental work examines the theft game in the absence of any detection mechanism. List (2007) and Bardsley (2008) study how the behavior typically observed in the dictator game changes once the possibility of theft is also introduced. Jakiela (2011) compares social norms of sharing within both a dictator game and a theft game across cultures.

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

possibility of type I errors) on the willingness to punish within a second-party punishment framework.[7]

Against this background, the present work contributes to the literature by focusing on the specific role played by type I errors for deterrence. In addition, we test experimentally the deterrence hypothesis within the setting of a theft game that has a very simple and intelligible structure and closely resembles a petty larceny. Our experimental design allows us to disentangle the incentive effect of adjudicative errors from other factors that may affect criminal behavior.

3. Theory

The basic model of crime deterrence considers an individual who has to decide whether to commit a crime. His initial wealth is w_0, and a successful crime makes him gain an amount equivalent to g in monetary terms. When the enforcement authority detects a criminal activity, it brings the evidence collected in front of the adjudicative authority to obtain a sanction with a monetary-equivalent value of f. The null hypothesis is that the defendant is innocent. There are two types of judicial errors: innocent individuals mistakenly judged guilty (type I errors) and guilty individuals mistakenly judged innocent (type II errors). We assume that, given the amount of incriminatory evidence produced by the authority and the burden of evidence necessary to reach a verdict of guilt, there exist a given probability of type I error, denoted ε_1, and a given probability of type II error, denoted ε_2.[8] If the burden of evidence is increased, more type II errors and fewer type I errors are produced. If, instead, the forensic technology improves, the same amount of incriminatory evidence gives a more precise signal of guilt and, thus, the probabilities of type I and type II errors decrease.

The individual knows that if he decides to commit the crime (C), there is a probability ε_2 of avoiding due punishment. Conversely, if he does not commit the crime (I), there is a probability ε_1 of being wrongfully punished. Under the assumption of risk neutrality, the individual decides whether to commit the crime by comparing the expected payoff obtained in the two scenarios. The expected costs of committing the crime are determined by the magnitude of the sanction (f) and the probability of being deservedly punished ($1 - \varepsilon_2$). The

[7] It should be observed that monitoring and sanctioning scofflaw behavior are routinely part of experiments in other streams of research. In many experimental designs, subjects may undertake an action that increases their payoff but makes them potentially subject to some form of punishment. Experiments involving a principal-agent relationship generally have these characteristics and focus, for example, on team production in labor and personnel economics (Falk and Fehr 2003; Backes-Gellner et al. 2008; Falk and Gächter 2008), corruption (Schulze and Frank 2003; Abbink 2006), and tax avoidance (Torgler 2002). Indeed, most experimental studies of issues related to punishment, enforcement, and deterrence use public-good settings (Tyran and Feld 2006; Galbiati and Vertova 2008; Bernasconi, Corazzini, and Marenzi 2010; Engel and Irlenbusch 2010; Grechenig, Nicklisch, and Thöni 2010). Along these lines, Falk and Fischbacher (2002) use a reverse public-good game to examine how social interaction affects the propensity to commit a crime.

[8] The production of errors has been modeled extensively in the literature (see, for example, Miceli 2009; Lando 2009; Feess and Wohlschlegel 2009; Rizzolli and Saraceno, forthcoming).

expected costs of not committing the crime are determined by ε_1 and f. The expected payoffs in the two cases are

$$E\pi^{C} = \varepsilon_2(w_0 + g) + (1 - \varepsilon_2)(w_0 + g - f)$$

and

$$E\pi^{I} = \varepsilon_1(w_0 - f) + (1 - \varepsilon_1)w_0.$$

The individual will not commit the crime if $E\pi^{C} < E\pi^{I}$. This is true if

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

Equation (1) indicates that, under the assumption of risk neutrality, ε_1 and ε_2 have the same impact on deterrence. While type II errors increase the expected payoff of committing the crime, type I errors decrease the expected payoff of not committing the crime. At the margin, one additional innocent person convicted is as detrimental to deterrence as one additional guilty person acquitted (Png 1986).[9]

Proposition 1. Under risk neutrality, type I and type II errors are equally detrimental to deterrence.

More generally, if we relax the assumption of risk neutrality, the expected utility from crime and innocence can be written as follows:

$$\mathrm{EU}^{C} = \varepsilon_2 U(w_0 + g) + (1 - \varepsilon_2)U(w_0 + g - f)$$

and

$$\mathrm{EU}^{I} = \varepsilon_1 U(w_0 - f) + (1 - \varepsilon_1)U(w_0).$$

An individual who maximizes expected utility is deterred from committing the crime if $\mathrm{EU}^{C} < \mathrm{EU}^{I}$, that is, if

$$\begin{aligned} &U(w_0) - U(w_0 + g - f) - \varepsilon_1[U(w_0) - U(w_0 - f)] \\ &- \varepsilon_2[U(w_0 + g) - U(w_0 + g - f)] > 0. \end{aligned} \tag{2}$$

As shown in Figure 1, if agents are risk averse, the concavity of the utility function implies that $U(w_0) - U(w_0 - f) > U(w_0 + g) - U(w_0 + g - f)$. As a consequence, type I errors are more detrimental to deterrence than type II errors.

Proposition 2. Under risk aversion, type I errors are more detrimental to deterrence than type II errors.

There are also nonexpected utility factors that may explain asymmetries in the effects of judicial errors on deterrence. First, consider loss aversion, a behavioral concept that can explain many phenomena that are inconsistent with

[9] 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 2008, sec. 15).

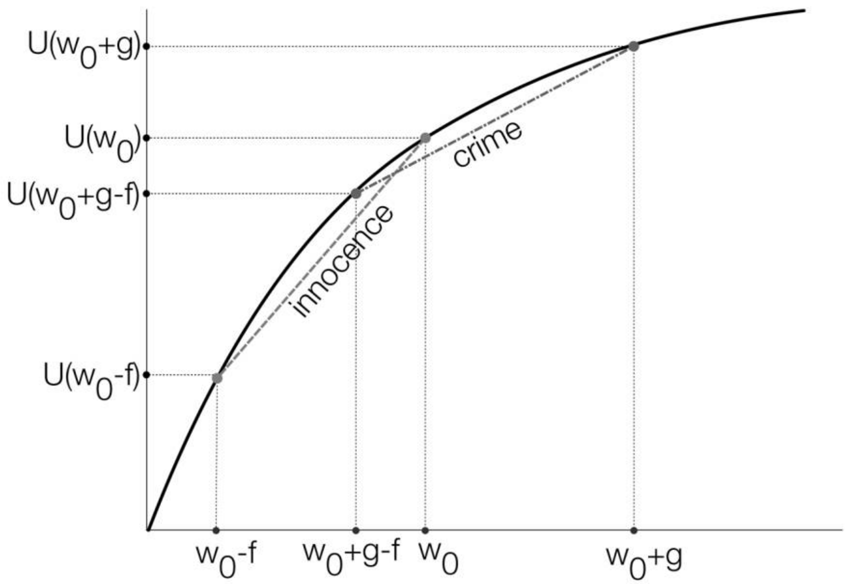

Figure 1. Effects of type I and type II errors on deterrence

the expected utility framework (Kahneman 2003; Kahneman, Knetsch, and Thaler 1990; Rabin 2000; Rabin and Thaler 2001).[10] Loss aversion is based on the premise that people tend to think of possible outcomes relative to a reference point, rather than in absolute terms, and that losses have a stronger psychological effect than corresponding gains. This may explain the tendency for people to prefer the avoidance of losses (outcomes below the reference point) to the acquisition of comparable gains (outcomes above the reference point). As shown by Kahneman and Tversky (1979), loss aversion can be represented with a utility function that is steeper for losses than for gains. Figure 2 displays the utility function with loss aversion (ULA) for the decision to commit a crime. The utility function has a kink at the reference point (w_0), the status quo before the crime. This implies that the utility function immediately to the left of the reference point is steeper than the one to the right.

In the presence of loss aversion, the deterrence condition of equation (2) can be written as follows:[11]

[10] Loss aversion was first proposed by Kahneman and Tversky (1979) in the framework of prospect theory and later developed for choice under uncertainty by Tversky and Kahneman (1991).

[11] Standard utility is commonly used in the literature to embody loss aversion (Schmidt and Zank 2005).

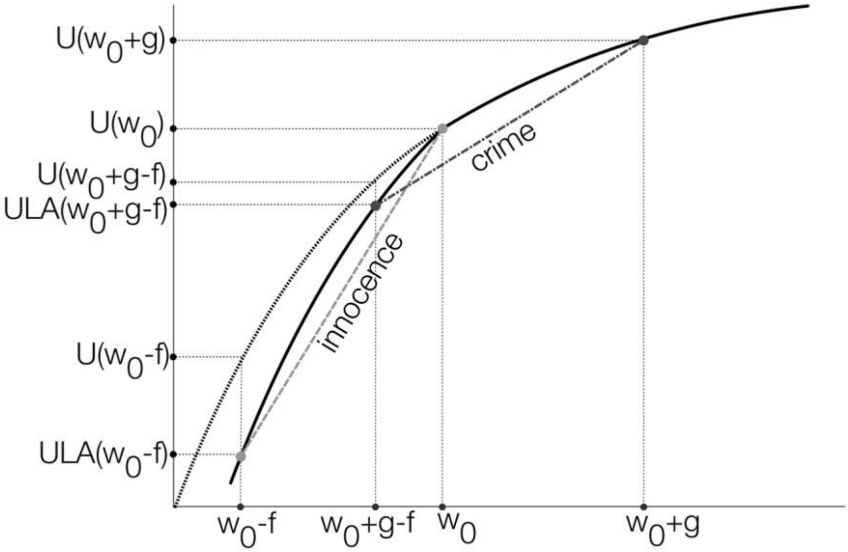

Figure 2. Effects of judicial errors on deterrence with loss aversion

$$\mathrm{ULA}(w_0) - \mathrm{ULA}(w_0 + g - f) - \varepsilon_1[\mathrm{ULA}(w_0) - \mathrm{ULA}(w_0 - f)]$$
$$- \varepsilon_2[\mathrm{ULA}(w_0 + g) - \mathrm{ULA}(w_0 + g - f)] > 0.$$

The asymmetry of the utility function at the reference point implies that $[\mathrm{ULA}(w_0) - \mathrm{ULA}(w_0 - f)] > [\mathrm{ULA}(w_0 + g) - \mathrm{ULA}(w_0 + g - f)]$; the innocent person wrongfully sanctioned suffers a loss that is larger than the corresponding gain for the guilty person wrongfully acquitted, since punishment represents a loss with respect to the status quo.

In addition to the effect of loss aversion, individuals may also be averse to wrongful convictions per se. Economists tend to have an instrumental view of the law as a set of incentives that constrain individuals only as long as it is optimal for them to abide. This approach, however, is not commonly shared, and most law scholars and philosophers emphasize the expressive function of rules:[12] the law prescribes a certain behavior, and people tend to follow its precepts because it is the right thing to do, with little regard for the sanction that backs the rule. Several studies indicate that individuals are generally inherently law-

[12] For economic approaches to the expressive function of law, see Cooter (1998). Indeed, not all economists have a Beckerian view of the law. Several microeconomic models assume that agents optimize subject to the constraint of the law (for some examples, see Nance 1997, p. 886 n.82). The assumption of unconstrained optimization has come under increasing criticism both theoretically (see Harrison 1986) and empirically (see, for example, Ellickson 1991).

abiding (Tyler 2006; Bicchieri 2006; Sacconi and Faillo 2010). They consider the law to be a guide for their behavior (Nance 1997) and gain a psychological payoff from conforming to rules.

Within this perspective, type I and type II errors are perceived differently. When a wrongful acquittal occurs, the violation of the prescribed behavior is not sanctioned, but at the same time the prescription is not questioned. With a wrongful conviction, in contrast, a certain behavior is first dictated and then reprimanded, so 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 averse to type I errors because they impose an additional cost in terms of loss of guidance and motivational crowding out. Summing up, for a given expected utility gain from crime, both loss aversion and type I error aversion imply that type I errors have a stronger impact on deterrence than do type II errors.

Proposition 3. If agents are loss averse or type I error averse, type I errors are more detrimental to deterrence than are type II errors.

4. Experimental Design

4.1. Baseline Game

The kind of crime mimicked in this experiment is petty larceny. The experimental task is based on a theft game, described in Figure 3.[13] Two agents, A and B, are randomly matched and assigned an initial endowment (w_A, w_B). Agent A has to decide whether to steal an amount, g, from B's endowment. If A decides to take g, there is a probability $1 - \varepsilon_2$ that the theft will be detected and, if this happens, A pays a sanction f (while keeping g). If A decides not to steal, he will be (wrongfully) sanctioned with probability ε_1. Note that A is not informed of the size of B's endowment. This allows us to abstract from issues related to distributional fairness; in Section 6 we discuss the possible effects of A's expectations about B's endowment. In the experimental task, the decision is therefore about the conformity to the social norm of not stealing. Within this setting, we can test whether and how type I and type II errors affect the willingness to abide by the rule.

[13] The theft game has received relatively little attention in the experimental literature, although it can be viewed as a reverse-dictator game. While in the dictator game the individual who owns the endowment decides about its allocation, in the theft game it is the individual who does not own the endowment who decides about its allocation.

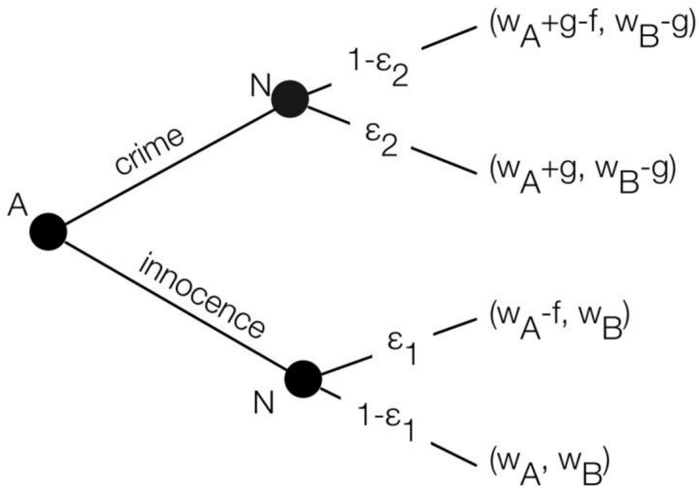

Figure 3. The theft game

4.2. Treatments

The experimental design is based on six treatments, as shown in Table 1. The endowment of A is $w_A = 10$ in treatments T1–T4 (all monetary values are expressed in euros). The endowment of B, unknown to A, is $w_B = 15$ in all treatments. The amount g that can be stolen from B and the fine f are both equal to 10 in all treatments.[14] The treatment variables are the probabilities of type I and type II errors (ε_1, ε_2). They are both set to zero in the control treatment (T1) that provides the benchmark under optimal deterrence. They are increased independently in T2 ($\varepsilon_1 = 0$; $\varepsilon_2 = .5$) and T3 ($\varepsilon_1 = .5$; $\varepsilon_2 = 0$) and jointly in T4 ($\varepsilon_1 = .25$; $\varepsilon_2 = .25$), while keeping constant the expected gain from crime. In T5 and T6 we replicate T2 and T3 while varying the endowment of A ($w_A = 0$ and $w_A = 20$ in T5 and T6, respectively), in order to compare the effects of judicial errors while keeping constant not only the expected gain but also the expected utility gain from crime.

Figure 4 provides a comparison of the six treatments. Note that, as shown in Table 1, the expected gain from committing the crime is the same in T2–T6 ($E\pi_A^C - E\pi_A^I = 5$). Also note that, relative to that in the control treatment (T1), the expected utility gain from crime is kept constant ($\Delta EU_A = [U(10) - U(0)]/2$) in T3 ($\varepsilon_1 = .5$; $\varepsilon_2 = 0$) and T5 ($\varepsilon_1 = 0$; $\varepsilon_2 = .5$). Likewise, the expected utility gain from crime is kept constant ($\Delta EU_A = [U(20) - U(10)]/2$) in T2 ($\varepsilon_1 = 0$; $\varepsilon_2 = .5$) and T6 ($\varepsilon_1 = .5$; $\varepsilon_2 = 0$).

[14] As the fine is set equal to the amount that can be potentially stolen, optimal deterrence can be achieved only in the absence of judicial errors. In all other cases, deterrence is suboptimal.

Table 1

Experimental Design: Comparison of Treatments

	T1	T2	T3	T4	T5	T6
ε_1	0	0	.5	.25	0	.5
ε_2	0	.5	0	.25	.5	0
w_A	10	10	10	10	0	20
$E\pi_A^I$	10	10	5	7.5	0	15
$E\pi_A^C$	10	15	10	12.5	5	20
EU_A^I	$U(10)$	$U(10)$	$[U(0) + U(10)]/2$	$[U(0) + 3U(10)[/4$	$U(0)$	$[U(10) + U(20)]/2$
EU_A^C	$U(10)$	$[U(10) + U(20)]/2$	$U(10)$	$[3U(10) + U(20)]/4$	$[U(0) + U(10)]/2$	$U(20)$
ΔEU_A	0	$[U(20) - U(10)]/2$	$[U(10) - U(0)]/2$	$[U(20) - U(0)]/4$	$[U(10) - U(0)]/2$	$[U(20) - U(10)]/2$

Note. ε_1 = probability of type I error; ε_2 = probability of type II error; w_A = endowment of subject A; $E\pi_A^I$ = A's expected payoff if innocent; $E\pi_A^C$ = A's expected payoff if criminal; EU_A^I = A's expected utility if innocent; EU_A^C = A's expected utility if criminal; ΔEU_A = net expected utility gain from committing the crime.

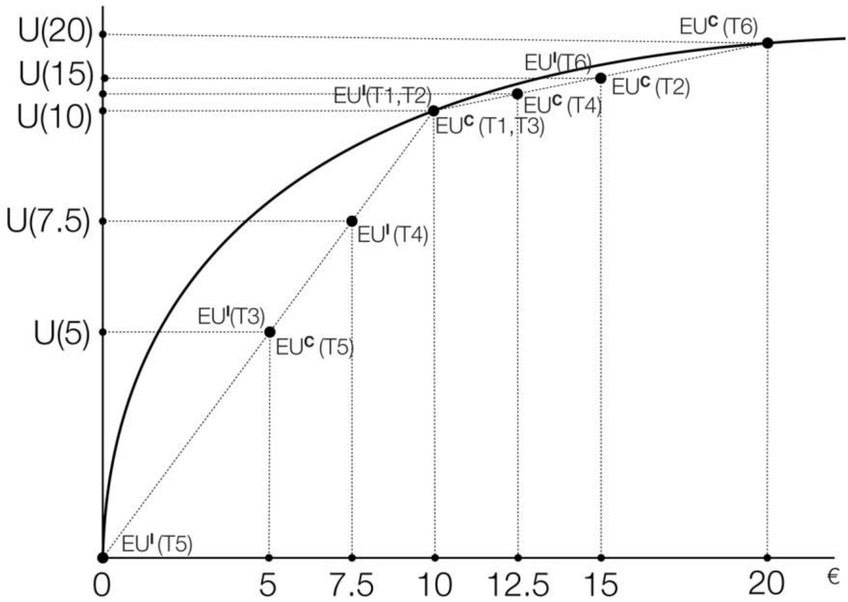

Figure 4. Comparison of expected utility across treatments

4.3. *Hypotheses*

Let us define Z_i as the fraction of agents in the population who opt for crime in treatment *i*. The first set of hypotheses concern the effects of judicial errors on deterrence.

Hypothesis 1a. A higher probability of type II errors leads to more crime.

The effect of wrongful acquittals on deterrence can be tested by comparing T1 ($\varepsilon_2 = 0$) with T2 ($\varepsilon_2 = .5$):

$$H_0 = Z_{T2} \leq Z_{T1} \quad \text{versus} \quad H_1 = Z_{T2} > Z_{T1}.$$

The next hypothesis is based on the Png (1986) extension of the deterrence hypothesis to type I errors.

Hypothesis 1b. A higher probability of type I errors leads to more crime.

The effect of wrongful convictions on deterrence can be tested by comparing T1 ($\varepsilon_1 = 0$) with T3 ($\varepsilon_1 = .5$):

$$H_0 = Z_{T3} \leq Z_{T1} \quad \text{versus} \quad H_1 = Z_{T3} > Z_{T1}.$$

Next we consider the effect of a joint increase in both type I and type II errors for a given expected gain from crime.

Hypothesis 1c. A higher probability of both type I and type II errors leads to more crime.

This hypothesis is tested by comparing T1 with T4:

$$H_0 = Z_{T1} \leq Z_{T4} \quad \text{versus} \quad H_1 = Z_{T4} > Z_{T1}.$$

The second set of questions we address is whether type I and type II errors have the same impact on deterrence. In the presence of risk aversion, loss aversion, or type I error aversion, we have hypothesis 2a.

Hypothesis 2a. Type I errors have a stronger effect on deterrence than type II errors.

We test this hypothesis by comparing T2 with T3. These two treatments differ only with respect to ε_1 and ε_2, while the expected monetary gains from crime are the same. Formally,

$$H_0 = Z_{T2} \geq Z_{T3} \quad \text{versus} \quad H_1 = Z_{T2} < Z_{T3}.$$

Next we focus on the specific role played by nonexpected utility factors, such as loss aversion or sensitivity to the expressive function of law. We test the effects of type I and type II errors on deterrence, while keeping constant the expected utility gain from crime, by comparing T2 with T6 and T3 with T5, respectively. Note that, as shown in Table 1, there is the same difference in expected utility between crime and honesty within each pair of treatments. Therefore, a stronger adverse effect of type I errors on deterrence would be consistent with either loss aversion or sensitivity to the expressive function of law.

Hypothesis 2b. For a given expected utility gain from crime, loss aversion or type I error aversion implies that type I errors have a stronger effect on deterrence than type II errors.

Formally, the null hypothesis can be stated as follows:

$$H_0 = Z_{T2} \geq Z_{T6} \quad \text{versus} \quad H_1 = Z_{T2} < Z_{T6}$$

and

$$H_0 = Z_{T3} \leq Z_{T5} \quad \text{versus} \quad H_1 = Z_{T3} > Z_{T5}.$$

4.4. Procedures

We implemented the treatments shown in Table 1 in two separate studies. Study 1 is based on a within-subjects design. Study 2 replicates the experiment in a between-subjects design, in order to rule out any order effects while turning the cold choices of the within-subjects design into a single hot choice (see Brandts and Charness 2000; Brosig, Weimann, and Yang 2003; Casari and Cason 2009).

In study 1 (sessions 1–2), T1 and T4 were played first, with a reverse order in each session. Treatments T2, T3, T5, and T6 were then played twice, in a switched sequence, thus providing an assessment of the role played by experience for the treatment effects. The theft game was therefore played by each subject repeatedly over 10 phases in a balanced crossover design.[15] Within each session, subjects were matched in pairs in each phase through a perfect-strangers 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 ex post each subject played only one role. The experiment employed a no-feedback design, so we could elicit the relevant responses from all subjects while obtaining statistically independent observations across subjects, although not within subjects, which we addressed by clustering standard errors at the individual level. Subjects were paid on the basis of the outcome of one randomly selected phase and one randomly selected role out of the two they had played.

Study 2 (sessions 3–6) replicated study 1 with a between-subjects design and a larger number of subjects. Note that since study 2 focused specifically on hypotheses 2a and 2b (differences in the effects of type I and type II errors), we implemented only T2, T3, T5, and T6, each with a different set of subjects. As in study 1, each subject played the role of player A, being informed that ex post each subject would play only one role. In each session of both studies, subjects were randomly assigned to computer terminals on their arrival. To ensure public knowledge, instructions were distributed and read aloud (see the Appendix). Sample questions were distributed to ensure understanding of the experimental task and procedures. Answers were privately checked and, if necessary, individually explained to the subjects, and the experiment did not start until all subjects had answered all questions correctly.

The experiments were conducted in the Experimental Economics Laboratory of the University of Milan–Bicocca between April 2009 (study 1) and May 2011 (study 2). None of the subjects had participated previously in similar theft games. We implemented six sessions, with 48 subjects in study 1 and 144 subjects in study 2, for a total of 192 subjects. Kruskal-Wallis tests indicated that there were no significant differences in the sociodemographic characteristics of the subjects across sessions. Participants were undergraduate students recruited by e-mail from a list of voluntary potential candidates. The average age and year of enrollment were 22.7 (range, 19–26; SD = 1.6) and 2.8 (range, 1–5; SD = 1.3), respectively. About 52 percent of the subjects had a nonvocational high school diploma. A university degree was reported as the highest educational attainment for father or mother by 21.9 and 11.5 percent of the subjects, respectively. No attendance fee was paid. Payments ranged between 0 and 20 euros, and the average payment was 11.1 (12.4) euros for sessions lasting on average approximately 45 (25) minutes in study 1 (study 2). The experiment was run with the experimental software z-Tree (Fischbacher 2007).

[15] The exact sequence of treatments was as follows: session 1: 1, 4, 2, 3, 5, 6, 2, 3, 5, 6; session 2: 4, 1, 3, 2, 6, 5, 3, 2, 6, 5.

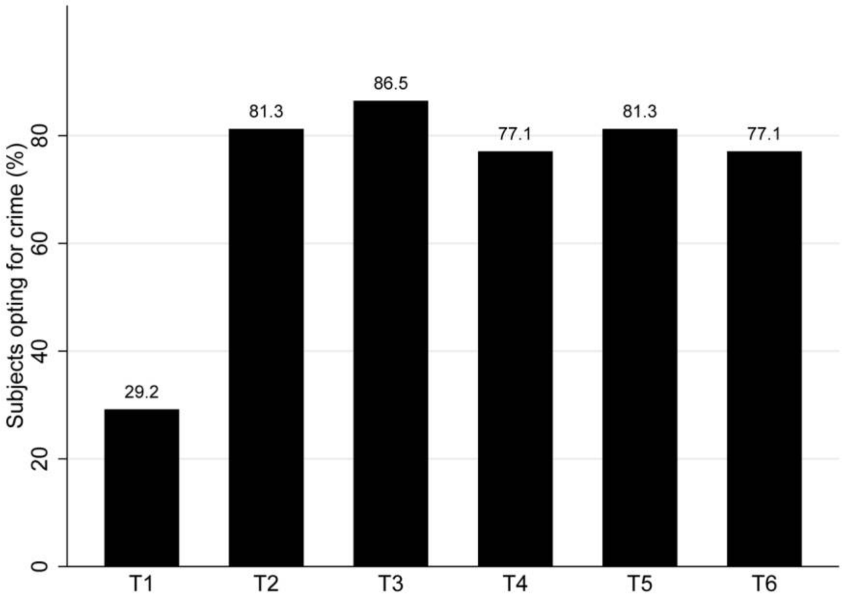

Figure 5. Percentage of subjects opting for crime, study 1

5. Results

5.1. Study 1

Figure 5 displays the percentage of subjects who decide to steal, by treatment, in study 1 (within-subjects design).[16] In T1, in which there is optimal deterrence ($\varepsilon_1 = \varepsilon_2 = 0$), the percentage of criminals is 29.2 percent. This figure provides the benchmark against which to assess the effects of judicial errors on deterrence. It should be observed that the percentage of criminals is exactly 29.2 percent in both sessions 1 and 2, despite the sequence of treatments being reversed. Similar results, irrespective of the sequence of treatments, were also obtained for T4 (75 and 79 percent in sessions 1 and 2, respectively; $p < .73$, Wilcoxon rank-sum test). This indicates that the order of treatments does not play a significant role.

The comparison between treatments indicates that judicial errors have a strong impact on deterrence. When ε_1 and ε_2 are increased jointly to .25 (T4), the percentage of criminals rises to 77.1. The effect is indeed stronger when, keeping constant the expected relative gains from crime, errors are increased individually. In T2, in which ε_2 is increased to .5, the crime rate rises to 81.3 percent. In T3, in which ε_1 is increased to .5, the crime rate rises to 86.5 percent. As an assessment

[16] For T2, T3, T5, and T6, the results refer to the two choices made by each subject (phases 3–6 and 7–10).

Table 2
Effects of Judicial Errors on Deterrence, Study 1

	T1, T2 (1)	T1, T3 (2)	T1, T4 (3)	T1, T2 (4)	T1, T3 (5)
T2 ($\varepsilon_2 = .5$; $\varepsilon_1 = 0$)	.52** (.08)			.47** (.09)	
T3 ($\varepsilon_1 = .5$; $\varepsilon_2 = 0$)		.57** (.07)			.62** (.08)
T4 ($\varepsilon_1 = \varepsilon_2 = .25$)			.48** (.09)		
T2 × phases 3–6				.11 (.10)	
T3 × phases 3–6					−.10 (.10)
N	144	144	96	144	144

Note. The dependent variable is the binary decision to steal. Probit estimates of marginal effects (expected change as the relevant dummy goes from 0 to 1) are reported. The omitted (reference) group is T1 (optimal deterrence, $\varepsilon_1 = \varepsilon_2 = 0$). Robust standard errors clustered by individual are in parentheses.
** $p < .01$.

of the statistical significance of these treatment effects, we report in Table 2 probit estimation results for the marginal effects of error probabilities on the decision to steal. We estimate three separate specifications for T2, T3, and T4. To assess the role played by repetition, we also consider, for T2 and T3, a specification that includes an interaction between the relevant treatment and a dummy for phases 3–6. Test statistics are based on robust standard errors clustered by individuals to account for within-subject dependence.

The effect of error probabilities on deterrence is strongly statistically significant in all three baseline specifications (columns 1–3). For a given expected gain from crime, the effect is stronger when judicial errors are increased individually rather than jointly. Columns 4 and 5 indicate that there is no evidence of order effects. The interaction between the relevant treatment variable and the dummy for phases 3–6 is not statistically significant ($p < .27$ and $p < .32$ for T2 and T3, respectively), which indicates that the effects of error probabilities are not related to repetition. The results are also qualitatively unchanged when we control for individual fixed effects: the estimated marginal effects are .75, .89, and .74 for T2, T3, and T4, respectively, and individually are strongly statistically significant ($p < .01$). Overall, these results indicate that both types of judicial errors have a large and significant effect on crime deterrence.

Result 1. A higher probability of either a type I or a type II error causes a significant increase in crime.

We now turn to hypotheses 2a and 2b: do type I and type II judicial errors have the same impact on deterrence? Table 3 presents probit estimation results for the differences in the effects of error probabilities. We start by considering crime rates in T2 and T3. The difference is positive (.05) and marginally sta-

Table 3
Differences in the Effects of Type I and Type II Errors, Study 1

	T2, T3 (1)	T2, T3, T5, T6 (2)	T2, T3, T5, T6 (3)
Type I error	.05^{+}	.01	−.04
	(.04)	(.04)	(.05)
Large EU gain		.05^{+}	−.00
		(.03)	(.05)
Type I error × large EU gain			.09^{+}
			(.05)
N	192	384	384

Note. The dependent variable is the binary decision to steal. Probit estimates of marginal effects (expected change as the dummy goes from 0 to 1) are reported. Robust standard errors clustered by individual are in parentheses. EU = expected utility.
$^{+}$ $p < .10$.

tistically significant ($p < .08$) for the relevant one-sided hypothesis. To control for differences in the expected utility gains from crime, we also include T5 and T6. Note that $\Delta EU_A = [U(20) - U(10)]/2$ in T2 and T6, and $\Delta EU_A = [U(10) - U(0)]/2$ in T3 and T5. We therefore obtain a 2 × 2 design in which two experimental conditions (error type and size of expected utility gain from crime) are varied independently. Column 2 shows the results of an examination of the main effects. The difference between the effects of type I and type II errors is small and not statistically significant ($p < .44$). The effect of the size of the expected utility gain is larger and marginally statistically significant ($p < .08$). The asymmetric effects of type I and type II errors appear to be entirely explained by differences in the expected utility gains from crime. Column 3 provides some evidence of an interaction between the experimental conditions: the difference between type I and type II error probabilities is positive and marginally significant only in the presence of a large expected utility gain (T3 and T5). As with the previous results, these results are qualitatively unchanged when we control for individual fixed effects.

Overall, the results of study 1 indicate that type I errors have a stronger effect on deterrence than type II errors, although the difference is only marginally significant. In addition, this difference can be explained within the expected utility framework. When the expected utility gain from crime is kept constant, there are no differences in the effects of type I and type II errors. To assess the robustness of these results, while ruling out any order and repetition effects, we replicate the experiment in study 2 using a between-subjects design with a single-shot decision and more subjects.

5.2. Study 2

Figure 6 displays the percentage of criminals by treatment in study 2. The hot-choice setting clearly makes the differences between treatments more salient: the percentage of subjects who opt for crime is 75.0 percent in T2 ($\varepsilon_2 = .5$;

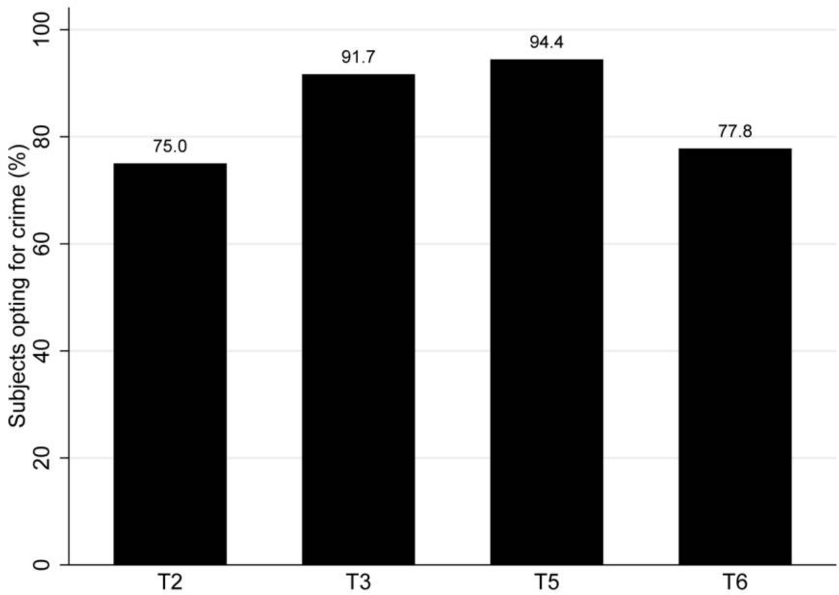

Figure 6. Percentage of subjects opting for crime, study 2

small ΔEU_A) and 91.7 percent in T3 ($\varepsilon_1 = .5$; large ΔEU_A). However, the control treatments indicate that this difference is not attributable to type I and type II errors per se but rather to the difference in the expected utility gains from crime: the percentage of subjects who choose to steal is 94.4 percent in T5 ($\varepsilon_2 = .5$; large ΔEU_A) and 77.8 percent in T6 ($\varepsilon_1 = .5$; small ΔEU_A).

To assess the statistical significance of these differences, Table 4 reports probit estimation results for the effects of type I errors, relative to type II errors, on the decision to steal. We also consider specifications that include controls for individual sociodemographic characteristics (age, gender, high school grade point average, and parental education). Focusing on T2 and T3 only, we see that the differential effect of type I errors is large and statistically significant ($p < .03$). This result is also robust when we control for subjects' sociodemographic characteristics ($p < .04$). Indeed, none of the individual characteristics are significantly related to the probability of committing the crime. Overall, the evidence is consistent with hypothesis 2a.

Result 2. Type I errors have a stronger impact on deterrence than type II errors.

Next we compare the effects of type I and type II errors while also including T5 and T6, thus keeping constant the expected utility gain from crime. The results from both specifications indicate that the difference between the effect

Table 4
Differences in the Effects of Type I and Type II Errors, Study 2

	T2, T3		T2, T3, T5, T6		T2, T3, T5, T6	
	(1)	(2)	(3)	(4)	(5)	(6)
Type I error	.17*	.13*	−.00	−.04	.02	−.01
	(.09)	(.08)	(.06)	(.06)	(.07)	(.07)
Large EU gain			.17**	.19**	.20*	.23**
			(.06)	(.06)	(.09)	(.09)
Type I error × large EU gain					−.07	−.09
					(.14)	(.15)
Age		.03		.03+		.03+
		(.02)		(.01)		(.01)
Gender (male)		.13		.04		.04
		(.08)		(.06)		(.06)
High school grade point average		.00		.00		.00
		(.00)		(.00)		(.00)
Parental education		.07		.03		.02
		(.09)		(.06)		(.06)
N	72	72	144	144	144	144

Note. The dependent variable is the binary decision to steal. Probit estimates of marginal effects (expected change as the dummy goes from 0 to 1) are reported. Robust standard errors are in parentheses. EU = expected utility.
+ $p < .10$.
* $p < .05$.
** $p < .01$.

of type I and type II errors is small and not statistically significant, whereas the effect of the size of the expected utility gain is large and strongly significant ($p < .01$). In addition, columns 5 and 6 indicate that there is no significant interaction between error type and the size of the expected utility gain from crime.

Result 3. After controlling for differences in the expected utility gain from crime, there is no significant difference in the effects of type I and type II errors.

Summing up, the results from study 2 indicate that type I errors have a stronger impact on deterrence than type II errors. Consistent with study 1, the asymmetric effects of type I and type II errors can be entirely explained within the expected utility framework. When the expected utility gains from crime are kept constant across treatments, there is no evidence of asymmetric effects produced by either loss aversion or type I error aversion.

6. Discussion

Our analysis indicates that, as expected, type I errors have a stronger impact on deterrence than type II errors. In contrast, there is no evidence of asymmetric effects once we control for differences in the expected utility gains from crime. Here we discuss additional issues that may be relevant for interpreting both of

Table 5

Crime and Payoff Inequality across Treatments

	T2	T3	T5	T6
(w_A, w_B)	(10, 10)	(10, 10)	(0, 0)	(20, 20)
$E\pi_A^I - \pi_B^I$	0	−5	0	−5
$E\pi_A^C - \pi_B^C$	15	10	15	10
$\lvert E\pi_A^C - \pi_B^C \rvert - \lvert E\pi_A^I - \pi_B^I \rvert$	15	5	15	5

Note. Hypothetical payoffs if A assumes $w_B = w_A$ are shown.

these results and, more generally, the impact of wrongful convictions on deterrence.

First, consider the role of distributive preferences. In our experimental task, subject A was not informed about the size of subject B's endowment. This choice was aimed at minimizing the confounding role of payoff inequality on the decision to commit the crime, in order to give more salience to the specific features of the two types of judicial errors. The underlying assumption is that, without information on the other subject's endowment, an individual does not take into account the distribution of payoffs when deciding whether to steal. However, it is possible that, in the absence of explicit information, subject A implicitly makes some assumption about B's endowment. In particular, one plausible scenario is that A assumes that B has an identical endowment.

Table 5 illustrates how inequality is affected by the decision to steal in T2, T3, T5, and T6 under the assumption that $w_B = w_A$.[17] The expected relative gain from crime is the same in all treatments: $(E\pi_A^C - \pi_B^C) - (E\pi_A^I - E\pi_B^I) = 15$. The effect of crime on absolute inequality $(\lvert E\pi_A^C - \pi_B^C \rvert - \lvert E\pi_A^I - \pi_B^I \rvert)$ is instead stronger in T2 than in either T3 or T6.[18] To assess the effects of inequality on the decision to steal under this scenario, we need to consider explicitly the specification of distributive preferences.

There are numerous attempts to incorporate inequity aversion into formal models of decision (Fehr and Schmidt 1999; Bolton and Ockenfels 2000; see also Binmore and Shaked 2010; Fehr and Schmidt 2010). These models generally share two main features: subjects dislike inequitable outcomes and suffer more from disadvantageous inequality than from advantageous inequality. Following Fehr and Schmidt (1999), we assume the following simplified description of player A's preferences:

$$U_A(\pi_A, \pi_B) = \pi_A - \delta_+(\pi_A - \pi_B) \quad \text{if } \pi_A > \pi_B \tag{3}$$

and

$$U_A(\pi_A, \pi_B) = \pi_A - \delta_-(\pi_B - \pi_A) \quad \text{if } \pi_B > \pi_A, \tag{4}$$

[17] Note that this assumption implies the possibility of a negative payoff in T5, where $w_A = 0$.

[18] If A does not steal, the difference in the expected payoff between A and B is 0 in T2 and −5 in both T3 and T6. If A opts for crime, the difference in the expected payoff is 15 in T2 and 10 in both T3 and T6.

where $\delta_+ \geq 0$ and $\delta_- \geq 0$ represent the weight of advantageous and disadvantageous inequality, respectively.

If agents place the same weight on advantageous and disadvantageous inequality ($\delta_+ = \delta_-$), we obtain

$$U_A(\pi_A, \pi_B) = \pi_A - \delta\,|\pi_B - \pi_A|\,, \tag{5}$$

so the effect of inequality on crime deterrence depends on $(|E\pi_A^C - \pi_B^C| - |E\pi_A^I - \pi_B^I|)$. Therefore, since crime increases inequality relatively more in T2, inequity aversion should imply relatively more crime under type I errors (T3 and T6) than under type II errors (T2 and T5). If agents value disadvantageous inequality more than advantageous inequality ($\delta_- > \delta_+$), this asymmetry is indeed reinforced. In the presence of type I errors, subjects can avoid disadvantageous inequality (-5) by committing the crime. On the contrary, in the presence of type II errors, subjects can only increase advantageous inequality by committing the crime. In sum, if subject A makes the default assumption that B has an identical endowment, inequity aversion can be expected to reinforce the asymmetry in the effects of type I and type II errors. This prediction, however, is not supported by our experimental results.

A second set of issues concerns how emotions, such as guilt and shame, may influence the decision to commit a crime in the presence of judicial errors.[20] One specific feature of wrongful convictions is that they involve a notion of public shame. As the defendant's peers can observe the conviction, they may impose a vast array of shaming sanctions on the individual.[21] Shame is thus reinforced when the defendant's conviction becomes common knowledge. The degree of publicity of the sanction is often manipulated by the enforcement authority in order to reinforce deterrence. This manipulation can be taken to the extreme case of purely shaming sanctions, such as signs or apologies, that impose on the defendant no cost other than the shamefulness of the act (Kahan and Posner 1999; Kahan 2006; Massaro 1997; Harel and Klement 2005).

Shame is therefore a powerful driver of law-abiding behavior, and its manipulation might have an asymmetric impact on the propensity to commit crime in our experimental treatments. In particular, increasing the publicity of the sanction might substantially decrease the crime rate under type II errors but should not affect the crime rate under type I errors. This is because, in the presence of type II errors, individuals can avoid shame by abstaining from crime,

[20] Behavioral and experimental economics have long considered the impact of guilt and shame on behavior (Elster 1998; de Hooge, Zeelenberg, and Breugelmans 2007; Kurzban, DeScioli, and O'Brien 2007; Battigalli and Dufwenberg 2007). Accordingly, the law and economics literature has also drawn normative implications from these findings (see, among others, McAdams and Rasmusen 2007; Kaplow and Shavell 2007). See also Mitchell (2002) for a critique.

[21] Elster (1998, p. 67) argues that "the material sanctions themselves are best understood as vehicles of the emotion of contempt, which is the direct trigger of shame. When a person refuses to deal with someone who has violated a social norm, the latter may suffer a financial loss. More important, he will see the sanction as a vehicle for the emotions of contempt or disgust, and suffer shame as a result."

whereas in the presence of type I errors, abstaining from committing the crime can still lead to a sanction and thus does not avoid shame. In our experiment, however, the no-feedback design implies that public shame plays no role. This might help to explain the finding of no difference in the effects of type I and type II errors when we controlled for the expected utility gain of crime. Future research will have to assess to what extent shame matters in explaining the asymmetric impact of errors on deterrence.

7. Conclusions

Judicial errors against innocent defendants represent an important issue not only theoretically but also empirically. Consider the case of mistaken convictions in criminal procedure. Although no country produces official statistics on judicial errors, there is evidence provided by various national and international organizations.[22] Gross et al. (2005) emphasize how most relevant data sets are based on exoneration cases:[23] their own database lists 340 exonerations in the United States between 1989 and 2003. But counting only those who eventually obtain an official exoneration, they argue, may be just 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" (Gross et al. 2005, p. 551). It should also be emphasized that, although our argument is mainly framed in the context of criminal law, judicial errors can occur in any adjudicative process. Our analysis can therefore be relevant for the decisions made by several types organizations, such as administrative agencies, commercial trade associations, religious bodies, and professional sports leagues.

The first aim of this paper was to extend the economic theory of crime deterrence to account for the role of type I errors. We argued that both risk aversion and loss aversion contribute to produce an asymmetry in the effects of the two types of judicial errors: wrongful convictions are expected to be more detrimental to deterrence than wrongful acquittals. Moreover, we formulated the hypothesis that this asymmetry can be reinforced for law-abiding individuals who are concerned about type I errors undermining the lawfulness of their behavior. The second aim was to provide an empirical test of the predictions of our extended model through a laboratory experiment. The experimental design was based on a theft game that simulated a simple crime, petty larceny. We

[22] The Web site Forejustice (http://www.forejustice.org), for instance, provides a list of 2,803 cases of innocents convicted (see also Fon and Schaefer 2007). This list includes cases from 84 countries. In these 2,803 cases, 184 people were executed, 514 people were sentenced to death, 575 people were sentenced to life in prison, 2,622 people were judicially exonerated or pardoned, and 414 people were exonerated after a false confession as of July 2009.

[23] Exoneration cases are those 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, are declared innocent because new evidence emerged.

exogenously manipulated the probability of type I and type II errors and compared across treatments the propensity of individuals to commit the crime.

Our findings indicate that wrongful convictions have a significant effect on deterrence. In addition, consistent with the theoretical predictions, for a given expected relative gain from crime, type I errors have a stronger impact on deterrence than type II errors. We also find that this asymmetry in the effects of judicial errors can be explained entirely within the expected utility framework. Nonexpected utility factors, such as loss aversion and type I error aversion, are not found to play a significant role. Further work will explore how the effects of judicial errors are influenced by social preferences and emotions such as shame.

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 has been predicted in the literature. This makes an economic case for the public authority to place particular emphasis on type I errors in order to achieve optimal deterrence. The results presented in this paper also have important implications for the law and economics literature. In particular, they provide an economic interpretation of the pro-defendant bias observed in the criminal procedures of modern democracies. More generally, they contribute to explaining the common wisdom that the conviction of an innocent individual should be considered to be far worse than the acquittal of a guilty individual.

Appendix

Instructions: Study 1

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

1. There are 24 subjects participating in this experiment.
2. The experiment takes place in 10 independent phases. Instructions for each phase will appear on the screen.
3. In each phase, 12 couples of two participants will be formed randomly and anonymously, so in each phase you will interact with a different subject.
4. Within each couple, the two subjects will be randomly assigned two different roles: A and B.
5. Therefore, in each phase each subject will interact exclusively with the other

subject in his/her pair, without knowing his/her identity, with the role (A or B) assigned with equal probability.

6. 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.

7. 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

1. Your choice consists of deciding whether to steal 10 euros from the subject you have been paired with.

2. You will have a known endowment of x euros, whereas the other subject will have an endowment of y euros (unknown to you).

3. After you have made your choice, there will be a check in order to punish those who steal.

4. The check will be subject to a certain error, as follows:
   a. If you have stolen 10 euros from the other subject, you will have to pay a 10-euro fine with a probability of p.
   b. If you have not stolen 10 euros from the other subject, you will have to pay a 10-euro fine with a probability of q.

5. Therefore, within each phase, earnings will be determined as follows:
   a. If you decide to steal, the other subject will earn $y - 10$ euros and you will earn
      x euros with a probability of p, or
      $x + 10$ euros with a probability of $1 - p$.
   b. If you decide not to steal, the other subject will earn y euros and you will earn
      $x - 10$ euros with a probability of q, or
      x euros with a probability of $1 - q$.

How Final Earnings Are Determined

1. 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.

2. 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.

References

Abbink, Klaus. 2006. Laboratory Experiments on Corruption. Pp. 418–38 in *International Handbook on the Economics of Corruption,* edited by Susan Rose-Ackerman. Cheltenham: Edward Elgar.

Backes-Gellner, Uschi, Donata Bessey, Kerstin Pull, and Simone N. Tuor. 2008. What Behavioural Economics Teaches Personnel Economics. Working Paper No. 77. University of Zurich, Institute for Strategy and Business Economics, Zurich.

Bardsley, Nicholas. 2008. Dictator Game Giving: Altruism or Artefact? *Experimental Economics* 11:122–33.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. Guilt in Games. *American Economic Review* 97:170–76.

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

Bernasconi, Michele, Luca Corazzini, and Anna Marenzi. 2010. Expressive Obligations in Public Good Games: Crowding-In and Crowding-Out Effects. Working Paper No. 4/2010. University of Venice, Department of Economics, Venice.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Binmore, Ken, and Avner Shaked. 2010. Experimental Economics: Where Next? *Journal of Economic Behavior and Organization* 73:87–100.

Block, Michael K., and Joseph Gregory Sidak. 1980. The Cost of Antitrust Deterrence: Why Not Hang a Price Fixer Now and Then? *Georgetown Law Journal* 68:1131–39.

Bolton, Gary E., and Axel Ockenfels. 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90:166–93.

Brandts, Jordi, and Gary Charness. 2000. Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games. *Experimental Economics* 2:227–38.

Brosig, Jeannette, Joachim Weimann, and Chun-Lei Yang. 2003. The Hot versus Cold Effect in a Simple Bargaining Experiment. *Experimental Economics* 6:75–90.

Casari, Marco, and Timothy N. Cason. 2009. The Strategy Method Lowers Measured Trustworthy Behavior. *Economics Letters* 103:157–59.

Cooter, Robert. 1998. Expressive Law and Economics. *Journal of Legal Studies* 27:585–608.

Craswell, Richard, and John E. Calfee. 1986. Deterrence and Uncertain Legal Standards. *Journal of Law, Economics, and Organization* 2:279–303.

Dacey, Raymond, and Kenneth S. Gallant. 1997. Crime Control and Harassment of the Innocent. *Journal of Criminal Justice* 25:325–34.

DeAngelo, Gregory J., and Gary Charness. 2009. Uncertainty and Deterrence: Experimental Evidence. Working paper. University of California, Department of Economics, Santa Barbara.

de Hooge, Ilona E., Marcel Zeelenberg, and Seger M. Breugelmans. 2007. Moral Sentiments and Cooperation: Differential Influences of Shame and Guilt. *Cognition and Emotion* 21:1025–42.

de La Bruyère, Jean. [1688] 1696. *Les caractères*. Dernière edition. Paris: E. Michallet.

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

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge, Mass.: Harvard University Press.

Elster, Jon. 1998. Emotions and Economic Theory. *Journal of Economic Literature* 36:47–74.

Engel, Christoph, and Bernd Irlenbusch. 2010. Turning the Lab into Jeremy Bentham's Panopticon: The Effect of Punishment on Offenders and Non-Offenders. Working Paper No. 2010/06. Max Planck Institute for Research on Collective Goods, Jena.

Enukashvili, Michael Xavier. 2010. Punishment and Voluntary Norm Abidance. Ph.D. thesis. Ruprecht-Karls-Universität Heidelberg.

Falk, Armin, and Ernst Fehr. 2003. Why Labour Market Experiments? *Labour Economics* 10:399–406.

Falk, Armin, and Uri Fischbacher. 2002. "Crime" in the Lab-Detecting Social Interaction. *European Economic Review* 46:859–69.

Falk, Armin, and Simon Gächter. 2008. *Experimental Labour Economics.* Pp. 156–59 in vol. 3 of *The New Palgrave Dictionary of Economics,* edited by Steven N. Durlaf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan.

Feess, Eberhard, and Ansgar Wohlschlegel. 2009. Why Higher Punishment May Reduce Deterrence. *Economics Letters* 104:69–71.

Fehr, Ernst, and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114:817–68.

Fehr, Ernst, and Klaus M. Schmidt. 2010. On Inequity Aversion: A Reply to Binmore and Shaked. *Journal of Economic Behavior and Organization* 73:101–8.

Fischbacher, Urs. 2007. z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics* 10:171–78.

Fon, Vincy, and Hans B. Schaefer. 2007. State Liability for Wrongful Conviction: Incentive Effects on Crime Levels. *Journal of Institutional and Theoretical Economics* 163:269–84.

Galbiati, Roberto, and Pietro Vertova. 2008. Obligations and Cooperative Behaviour in Public Good Games. *Games and Economic Behavior* 64:146–70.

Garoupa, Nuno. 1997. The Theory of Optimal Law Enforcement. *Journal of Economic Surveys* 11:267–95.

Garoupa, Nuno, and Matteo Rizzolli. 2012. Wrongful Convictions Do Lower Deterrence. *Journal of Institutional and Theoretical Economics* 168:224–31.

Gould, Jon B., and Richard A. Leo. 2010. One Hundred Years Later: Wrongful Convictions after a Century of Research. *Journal of Criminal Law and Criminology* 100:825–68.

Grechenig, Kristoffel, Andreas Nicklisch, and Christian Thöni. 2010. Punishment despite Reasonable Doubt—a Public Goods Experiment with Sanctions under Uncertainty. *Journal of Empirical Legal Studies* 7:847–67.

Gross, Samuel R., Kristen Jacoby, Daniel J. Matheson, Nicholas Montgomery, and Sujata Patil. 2005. Exonerations in the United States, 1989 through 2003. *Journal of Criminal Law and Criminology* 95:523–60.

Gross, Samuel R., and Barbara O'Brien. 2008. Frequency and Predictors of False Conviction: Why We Know So Little and New Data on Capital Cases. *Journal of Empirical Legal Studies* 5:927–62.

Harbaugh, William T., Naci H. Mocan, and Michael S. Visser. 2011. Theft and Deterrence. Working Paper No. 17059. National Bureau of Economic Research, Cambridge, Mass.

Harel, Alon, and Alon Klement. 2005. The Economics of Shame: Why More Shaming May Deter Less. Discussion Paper No. 401. Hebrew University, Center for Rationality and Interactive Decision Theory, Jerusalem.

Harris, John R. 1970. On the Economics of Law and Order. *Journal of Political Economy* 78:165–74.

Harrison, Jeffrey L. 1986. Egoism, Altruism, and Market Illusions: The Limits of Law and Economics. *UCLA Law Review* 33:1309–54.

Immordino, Giovanni, and Michele Polo. 2008. Judicial Errors and Innovative Activity. Working Paper No. 196. Innocenzo Gasparini Institute for Economic Research, Milan.

Jakiela, Pamela. 2011. Social Preferences and Fairness Norms as Informal Institutions: Experimental Evidence. *American Economic Review* 101:509–13.

Kahan, Dan M. 2006. What's Really Wrong with Shaming Sanctions. *Texas Law Review* 84:2075–96.

Kahan, Dan M., and Eric A. Posner. 1999. Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines. *Journal of Law and Economics* 42: 365–91.

Kahneman, Daniel J. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review* 93:1449–75.

Kahneman, Daniel J., Jack L. Knetsch, and Richard H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98: 1325–48.

Kahneman, Daniel J., and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263–92.

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

———. 1998. Accuracy in Adjudication. Pp. 1–7 in *The New Palgrave Dictionary of Economics and the Law,* edited by Peter Newman. London: Macmillan.

Kaplow, Louis, and Steven Shavell. 1994. Accuracy in the Determination of Liability. *Journal of Law and Economics* 37:1–15.

———. 2007. Moral Rules, the Moral Sentiments, and Behavior: Toward a Theory of an Optimal Moral System. *Journal of Political Economy* 115:494–514.

Kurzban, Robert, Peter DeScioli, and Erin O'Brien. 2007. Audience Effects on Moralistic Punishment. *Evolution and Human Behavior* 28:75–84.

Lando, Henrik. 2006. Does Wrongful Conviction Lower Deterrence? *Journal of Legal Studies* 35:327–38.

———. 2009. Prevention of Crime and the Optimal Standard of Proof in Criminal Law. *Review of Law and Economics* 5:33–52.

Levitt, Steven D., and Thomas J. Miles. 2007. Empirical Study of Criminal Punishment. Pp. 455–95 in vol. 1 of *Handbook of Law and Economics,* edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: Elsevier.

List, John A. 2007. On the Interpretation of Giving in Dictator Games. *Journal of Political Economy* 115:482–93.

Massaro, Toni M. 1997. The Meanings of Shame: Implications for Legal Reform. *Psychology, Public Policy, and Law* 3:645–704.

McAdams, Richard H., and Eric B. Rasmusen. 2007. Norms and the Law. Pp. 1573–1618 in vol. 2 of *Handbook of Law and Economics,* edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: Elsevier.

Miceli, Thomas J. 2009. Criminal Procedure. Pp. 125–44 in *Criminal Law and Economics: Encyclopedia of Law and Economics,* edited by Nuno Garoupa. Cheltenham: Edward Elgar.

Mitchell, Gregory. 2002. Why Law and Economics' Perfect Rationality Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence. *Georgetown Law Journal* 91:67–167.

Nance, Dale A. 1997. Guidance Rules and Enforcement Rules: A Better View of the Cathedral. *Virginia Law Review* 83:837–937.

Png, Ivan P. L. 1986. Optimal Subsidies and Damages in the Presence of Judicial Error. *International Review of Law and Economics* 6:101–5.

Polinsky, A. Mitchell, and Steven Shavell. 2008. Law, Public Enforcement of. Pp. 38–47 in vol. 5 of *The New Palgrave Dictionary of Economics,*edited by Steven N. Durlauf and Lawrence E. Blume. 2d ed. Baskingstoke: Palgrave Macmillan

Rabin, Matthew. 2000. Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica* 68:1281–92.

Rabin, Matthew, and Richard H. Thaler. 2001. Risk Aversion. *Journal of Economic Perspectives* 15:219–32.

Rauhut, Heiko. 2009. Higher Punishment, Less Control? Experimental Evidence on the Inspection Game. *Rationality and Society* 21:359–92.

Rizzolli, Matteo, and Margherita Saraceno. Forthcoming. Better That Ten Guilty Persons Escape: Punishment Costs Explain the Standard of Evidence. *Public Choice.*

Sacconi, Lorenzo, and Marco Faillo. 2010. Conformity, Reciprocity, and the Sense of Justice: How Social Contract–Based Preferences and Beliefs Explain Norm Compliance: The Experimental Evidence. *Constitutional Political Economy* 21:171–201.

Schildberg-Hörisch, Hannah, and Christina Strassmair. 2012. An Experimental Test of the Deterrence Hypothesis. *Journal of Law, Economics, and Organization* 28:447–59.

Schmidt, Ulrich, and Horst Zank. 2005. What Is Loss Aversion? *Journal of Risk and Uncertainty* 30:157–67.

Schulze, Günter G., and Björn Frank. 2003. Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility. *Economics of Governance* 4:143–60.

Smith, Bruce P. 2004. The History of Wrongful Execution. *Hastings Law Journal* 56:1185–1234.

Sonnemans, Joep, and Frans van Dijk. Forthcoming. Errors in Judicial Decisions: Experimental Results. *Journal of Law, Economics, and Organization.*

Strandburg, Katherine. 2003. Deterrence and the Conviction of Innocents. *Connecticut Law Review* 35:1321–50.

Torgler, Benno. 2002. Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments. *Journal of Economic Surveys* 16:657–83.

Tversky, Amos, and Daniel J. Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference Dependent Model. *Quarterly Journal of Economics* 106:1039–61.

Tyler, Tom R. 2006. *Why People Obey the Law.* Princeton, N.J.: Princeton University Press.

Tyran, Jean-Robert, and Lars P. Feld. 2006. Achieving Compliance When Legal Sanctions Are Non-Deterrent. *Scandinavian Journal of Economics* 108:135–56.