
Munich Personal RePEc Archive

The Impact of Fine Size and Uncertainty on Punishment and Deterrence: Evidence from the Laboratory

Feess, Eberhard and Schramm, Markus and Wohlschlegel, Ansgar

28 September 2014

Online at https://mpra.ub.uni-muenchen.de/59463/
MPRA Paper No. 59463, posted 29 Oct 2014 15:11 UTC

The Impact of Fine Size and Uncertainty on Punishment and Deterrence: Evidence from the Laboratory

Eberhard Feess*, Markus Schramm† and Ansgar Wohlschlegel‡

September, 28th 2014

Abstract

We use a laboratory experiment to test the impacts of uncertainty, the magnitude of fines and aversion against making type-I and type-II errors on legal decision making. Measuring uncertainty as the noise of a signal on the defendant's guilt observed by legal decision makers, we observe that a supposed wrongdoer is less likely to be punished if fines and uncertainty are high. Furthermore, judges care far more about type-I errors and violators steal far less often than expected payoff maximizers would. While our results support the theoretical predictions on average, a cluster analysis provides evidence for heterogenous behavior of participants, many of whom don't respond to changes in the parameters or are far more driven by uncertainty than the magnitude of fines.

Keywords: Deterrence, fine size, type-I and type-II error, experiment

JEL codes: C91, D03, K14

*Frankfurt School of Finance & Management, Sonnemannstr. 9-11, 60314 Frankfurt, Germany, E-mail: e.feess@frankfurt-school.de

†RWTH Aachen University, Templergraben 64, 52056 Aachen, Germany. E-mail: markus.schramm1@rwth-aachen.de

‡Portsmouth Business School, Richmond Building Portland Street, Portsmouth PO1 3DE, United Kingdom, E-mail: Ansgar.Wohlschlegel@port.ac.uk

1 Introduction

Two fundamental insights of the economic analysis of law are that deterrence is increasing in the magnitude of fines (Becker, 1968) and the accuracy of the court system (Png, 1986, Polinsky and Shavell, 1999). For the latter point, it has been noted that both higher frequencies of convicting innocent defendants (type-I errors) and higher percentages of releasing guilty defendants (type-II errors) have detrimental effects as the incentives to obey the law depend on the difference in the expected fine with and without violating the law.

The deterrence effect of higher fines is straightforward when type-I and type-II errors are exogenously given, which is usually assumed in the literature.[1] In reality, however, the frequencies of these two error types do not only depend on the evidence and the accuracy of the court system, but also on the relative weights judges and juries put on them: For a given quality of evidence, which is a noisy signal on the suspects' actual behavior, legal decision makers can reduce type-I errors at the expense of higher type-II errors and vice versa. Intuitively, one would expect that the legal decision makers' aversion against making type-I errors increases in the fine size, and if this effect is strong enough, then higher *actual* punishment may even decrease *expected* punishment (Andreoni, 1991). Understanding the impact of the fine size on the relative frequency of type-I and type-II errors is hence a crucial point for the proper design of legal punishment.

Based on a model that structures the potential effects at work, we perform a laboratory experiment to analyze the interplay of legal uncertainty, represented by the precision of a noisy signal on the actual behavior, and the magnitude of fines in impacting on punishment and violation. Participants are divided into two groups, potential violators and judges. Both groups are informed that there is a fixed amount of money supposed to be donated to charity. The money can be stolen by the violators. If it is not stolen, it may still disappear due to a random event, and this creates legal uncertainty. Judges can punish if and only if they observe that the money has disappeared, and we vary both the level of fines and the degree of uncertainty. Note that the frequencies of type-I and type-II errors are endogenously determined by the judges' and potential violators' decisions, so that their interdependency is fully captured by our experiment.

Turning to our results, let us start with judges. First, we indeed find that the frequency of type-I errors relative to type-II errors decreases in the magnitude of fines. Thus, there is a countervailing effect on the

[1] See the overview in Polinsky and Shavell (2009), section 15.

deterrence effect of large fines, caused by the decision makers' lower willingness to accept type-I errors. Second, it is important to understand the relative weight judges put on type-I and type-II errors. This requires to take the actual consequences of misjudgments into account. We do so by using the fines imposed on innocents as a measure of the preference cost of making type-I errors, and the amount stolen by an unpunished violator as a measure of the preference cost of type-II errors. Comparing these amounts, we find that judges care far more about type-I than type-II errors, that is, they are more concerned about unjustified fines than about unatoned thefts.

For potential violators, we first find that the signal's noise turns out to be even more important than the magnitude of fines. Thus, our experiment reinforces the view that accurate decisions are crucial for the incentives to obey the law. Second, violators care more about their own money than about the money for donation, but they have (partially) social preferences or are averse against violating social norms. Given the punishment behavior of judges, stealing the money increases the expected payoff of potential violators for all of our combinations of fines and uncertainty, so that risk-neutral participants who do not care about the donation should always steal. This is contrasted by an overall stealing rate of around 55% in our experiment.

In summary, the participants' behaviour that we observe on average confirms our theoretical model's results. However, individual behaviour is very heterogenous. Indeed, performing a cluster analysis yields some additional insights into it: For instance, about 25% of judges and thieves respond far more to changes in uncertainty than to changes in fines. Moreover, another 25% of them don't respond at all to changes in fines and uncertainty. However, a closer look at these latter clusters reveals that 20% of judges never punish while only 4% punish in all cases, which is consistent with a far stronger aversion against type-I errors than against type-II errors.

Most of the experimental literature on deterrence has restricted attention to situations where not only the fines, but also the punishment probabilities are exogenously given. DeAngelo and Charness (2012) design an experiment where the expected fine is kept constant, while the probability of being punished and the uncertainty on this probability varies. They find that higher uncertainty enhances deterrence. Schildberg-Hörisch and Strassmair (2012) find that small sanctions reduce deterrence compared to no sanctions, which can be attributed to a crowding out of intrinsic motivation for socially appreciated behavior. While some individ-

uals act selfishly and reduce their criminal rate even for small sanctions, the majority of the participants responds in the expected direction only for large sanctions. Khadjavi (2014) confirms a path-dependency of the impact of fines known from experiments on rewards: When fines have a deterrence effect and are removed later on, then the pro-social behavior is lower compared to a situation with no fines at all. By using questionnaires on the participants' feelings, they can explain the observed behavior by a change in emotions. Rizzolli and Stanca (2012) find in the laboratory that, for identical incentives in case of selfish preferences, type I-errors have a larger adverse effect on deterrence than type II-errors.

The experimental papers on punishments which emerge endogenously from the behavior of the participants have adopted voluntary contribution mechanisms (VCMs) in which the participants can mutually sanction non-cooperative behavior (Fehr and Gächter, 2000). For our research question, however, a different setting seems appropriate: First, legal fines are better resembled by third-party punishment, where those who can punish are not directly affected by the behavior of violators.[2] Second, we assume that fines are costless for those who impose them, since judges themselves do not bear the social costs of punishment. Third, VCM games are, by definition, about voluntary payments, whereas we frame our experiment as a legal infringement by denoting the taking of the money provided for donation explicitly as theft.[3]

There are only a few VCM experiments with punishment which assume noisy signals or vary the magnitude of fines. Grechening et al. (2010) assume that participants get noisy signals on their mutual contributions in a VCM game and find that higher noise, which increases the risk of type-I errors in case of penalties, does not reduce the punishment frequency. Social welfare, however, shrinks due to retaliation of those who are innocently punished and because fines are costly.

While fines are kept constant in Grechening et al. (2010), Ambrus and Greiner (2012) consider different fine levels. With low fines, people do not punish often, thereby saving on punishment costs. When fines are high, then there is a large deterrence effect which improves social efficiency. As both of these beneficial effects are small for intermediate levels of punishment, efficiency is U-shaped in the magnitude of fines. This is related to results in Nikiforakis and Normann (2008) and Egas and Riedl (2008), which seem to be the

[2] One of the few papes applying third-party punishment is Fehr and Fischbacher (2004). but they assume that the behavior can be perfectly observed.

[3] While Rega and Telle (2004) find that a terminology that relates contributions to social norms has no impact on behavior, this may be different for dictions related to stealing.

first experiments on the impact of fine sizes on punishments and contributions in VCM settings. We are not aware of any other paper in which probabilities for type-I and II-errors are determined endogenously by the violators' actual behavior.

Other interesting results of VCM experiments that are, however, not directly related to variations in the noisy signal or the magnitude of sanctions include that violations of social norms are reinforcing (Falk and Fischbacher, 2002), that group decisions on punishments yield higher contributions (Ertan et al., 2009, and Casari and Luini, 2009), that revealing the identities of non-contributors increases cooperation (Masclet et al., 2003, and Rega and Telle, 2004) and that the deterrence effect of punishment declines sharply when counter-punishments are feasible (Nikiforakis, 2008). Bornstein and Weisel (2010) show that uncertainty about the counterparts' endowments reduces the benefits from the punishment option in a repeated VCM setting.

From a legal perspective, our finding that participants in the role of judges put considerably more weight on type-I errors compared to type-II errors is related to the famous Blackstone ratio that it is "better that ten guilty persons escape, than that one innocent suffers". Most people share the view that avoiding type-I errors matters most, but the marginal rate of substitution differs substantially among them (Volokh, 1997). To account for the Blackstone ratio, some models simply put higher weight on type-I errors in analyzing optimal judgements (Miceli, 1991, Lando, 2006), and several papers explain from a rational choice perspective why type-I errors are more severe for society (see Hylton and Khanna, 2007, who take up a public-choice perspective, and Persson and Siven, 2007, who adopt a median-voter model). By using a reversed dictator game where participants can steel from their counterparts' endowments, Rizzolli and Saraceno (2013) show experimentally that the adverse effects of type-I errors on deterrence are higher than those of type-II errors.

Finally, the starting point of our paper that higher fines may even reduce deterrence dates back to the early legal literature on nullification pioneered by Michael and Wechsler (1937). This literature argues that jurors or witnesses may not be willing to participate in legal action when punishments seem unfairly high, and theoretical models show that higher fines reduce deterrence when the willingness to accept type-I errors is decreasing in fines to a sufficiently large extent (Andreoni, 1991, Feess and Wohlschlegel, 2009).

The remainder of the paper is organized as follows: Section 2 provides a theoretical model on the

interdependency of the judges' and the potential violators' decisions. We show that the impacts of the signal's noise and the magnitude of fines on deterrence, and on the frequency of type-I and type-II errors in equilibrium, is not straightforward. In section 3, we describe the experimental setting. Section 4 provides descriptive statistics, and section 5 extends to regression analysis. Section 6 discusses the heterogeneity in the participants' behavior. We conclude in section 7.

2 A simple model of punishment and deterrence

We first develop a model that allows to analyze the impact of fine size and uncertainty on the interplay between punishment and deterrence. In basic models on errors in court (Png, 1986, Polinsky and Shavell, 1999), it is found that both type-I and type-II errors increase the violation frequency as the difference between the probability of being punished with and without violation shrinks. A full-fledged equilibrium analysis, however, needs to take into account that higher uncertainty and higher fines may reduce the decision makers' (judges or juries) willingness to punish, so that the relative probabilities for the two errors need to be derived endogenously.

The interdependency of the behavior of judges and violators implies that we cannot treat type-I and type-II errors as exogenous - the probability of convicting an innocent (type-I error) depends on the percentage of violators in equilibrium, which in turn depends on the (anticipated) punishment behavior of judges. Consider a potential violator who, in the case of an infringement, causes an adverse outcome (a 'loss') denoted by L.[4] In case of no infringement, an exogenous event causes the same loss L with probability q. Thus, we assume that the loss can occur only once; if the money is already stolen by the violator, it can't disappear for exogenous reasons any more.[5] In our experiment, observing L means that the money has disappeared, and q is the probability that this happens even in cases where it is not taken by the respective participant.

With ϕ as the percentage of participants who actually commit the act, the ex post probability that a violation took place after observing the loss is $\frac{\phi}{\phi+(1-\phi)q}$, and the ex post probability of no infringement is $\frac{(1-\phi)q}{\phi+(1-\phi)q}$. Thus, $\pi = \frac{(1-\phi)q}{\phi+(1-\phi)q}$ is the probability of committing a type-I error in case of punishment, while

[4] In our experimental design, this is the amount donated to charity.

[5] As in all models on errors in court, we need to assume that the facts of the case cannot be fully reconstrcuted ex post.

$1-\pi = \frac{\phi}{\phi+(1-\phi)q}$ is the probability of a type-II error in case of no punishment.

For judges, we introduce the following assumptions: First, we set the utility from correct decisions to zero, that is, we take only the preference costs of misjudgments, but not the benefits of correct decisions into account. Second, we assume that the disutility from type-I errors is $\alpha_i F$ where α_i is a parameter on judge i's aversion against type-I errors, and F is the fine size. Third, the disutility from type-II errors is $\beta_i A$ where β_i is the degree of judge i's aversion against type-II errors, and A the severity of the infringement which we will refer to as the "amount stolen". Normalizing $\beta_i = 1$, α_i captures the ratio of the degrees of aversion against type-I and type-II errors.

Recalling the probabilities of type-I and type-II errors, it follows that a judge i *who assumes a violation frequency* ϕ prefers to convict a suspect if and only if

$$\frac{(1-\phi)\,q}{\phi+(1-\phi)\,q}\alpha_i F \leq \frac{\phi}{\phi+(1-\phi)\,q}A. \tag{1}$$

Defining $\widetilde{\alpha}$ as the threshold type such that a judge prefers punishment for all $\alpha \leq \widetilde{\alpha}$, we get

$$\widetilde{\alpha} = \frac{A\phi}{Fq\,(1-\phi)}. \tag{2}$$

For potential violators, we define m_j as the weight that a potential violator j puts on the victim's payoff relative to her own payoff. For $m_j = 0$, violator j is completely selfish, and for $m_j = 1$, she puts equal weight on the victim's and her own payoff.

If individual j assumes punishment frequency p, she steals if and only if

$$A - pF \geq m_j A - qpF\text{:} \tag{3}$$

On the left hand side, the expected benefit is the difference in the amount stolen and the expected fine. On the right hand side, $m_j A$ is the utility associated with the donation,[6] and qp is the probability of being

[6]Identically, we could express $A\,(1-m_i)$ as the utility from stealing amount A where m_i captures the degree of disutility from violating a social norm.

punished by mistake (type-I error). Thus, we get

$$\widetilde{m} = 1 - \frac{Fp(1-q)}{A} \tag{4}$$

as threshold such that individual j violates if and only if $m_j \leq \widetilde{m}$.

Equations (2) and (4) characterize the judges' and violators' optimal decisions based on their expectations on their counterparts' behavior: Inspecting these two equations yields the following Proposition:

Proposition 1 *Suppose that violators treat the punishment frequency p, and judges the violation frequency ϕ, as exogenously given. Then:*

(i) $\frac{\partial \widetilde{m}}{\partial p} = -\frac{F(1-q)}{A} < 0$, $\frac{\partial \widetilde{m}}{\partial F} = -\frac{p(1-q)}{A} < 0$, *and* $\frac{\partial \widetilde{m}}{\partial q} = \frac{Fp}{A} > 0$ *and*

(ii) $\frac{\partial \widetilde{\alpha}}{\partial \phi} = \frac{A}{Fq(1-\phi)^2} > 0$, $\frac{\partial \widetilde{\alpha}}{\partial F} = -\frac{A\phi}{F^2 q(1-\phi)} < 0$, *and* $\frac{\partial \widetilde{\alpha}}{\partial q} = -\frac{A\phi}{Fq^2(1-\phi)} < 0$.

Part (i) of Proposition 1 first confirms the well-known deterrence theory: Violators are less likely to violate if they anticipate a higher punishment probability p and a larger fine size F. Furthermore, the violation frequency increases in q, that is, in the probability that the loss may also be observed without infringement. This resembles the literature on the impacts of errors in court discussed in the introduction.

Part (ii) discusses judges' behavioral responses to the model parameters when they seek to minimize expected preference costs from wrong decisions: When they anticipate violators to break the law more frequently, when the evidence on the actual punishment is less noisy, and when fines are low, then judges will also punish for higher aversion against type-I errors, so that $\widetilde{\alpha}$ increases. In a way, $\frac{\partial \widetilde{m}}{\partial p}$ characterizes how violators' best responses depend on the judges' punishment frequency as, for any given p, a potential violator steals if and only if $m \leq \widetilde{m}$. Similarly, $\frac{\partial \widetilde{\alpha}}{\partial \phi}$ captures the effect of the anticipated violating frequency on judges' best responses. While these individual best responses are intuitive, things are more involved when we take the interdependency of the decisions into account. In Bayesian Nash equilibrium, judges' and violators' expectations about their counterparts' actions coincide with their actual equilibrium choices, i.e.,

punishment and violation frequencies are given by the system of equations

$$p = G(\widetilde{\alpha}) \tag{5}$$

$$\phi = H(\widetilde{m}), \tag{6}$$

where $G(.)$ and $H(.)$ denote the cumulative distribution functions of judges' parameter α of relative aversion against type-I errors and potential violators' marginal rate of substitution m between the victim's and their own monetary payoff, provided that the thresholds $\widetilde{\alpha}$ and $\widetilde{m}$ are in the supports S_α of $G(.)$ and S_m of $H(.)$, respectively, the intersection of which determines the Bayesian Nash equilibrium.

The following proposition summarizes the impact of noise and the fine size on equilibrium punishment and violation frequencies, where judges and violators correctly anticipate their counterparts' equilibrium choices:

Proposition 2 *Suppose that S_α and S_m are intervals. If $\widetilde{\alpha} \in S_\alpha$ and $\widetilde{m} \in S_m$, then the Bayesian Nash equilibrium has the following comparative static properties:*

(i) For all distributions $G(\cdot)$ and $H(\cdot)$, $\frac{dp^*}{dF} < 0$ and $\frac{d\phi^*}{dq} > 0$.

(ii) Independently of the distribution $H(\cdot)$, $\frac{d\phi^*}{dF}$ is negative (positive, zero) if $G(\cdot)$ is concave (convex, linear).

(iii) $\frac{dp^*}{dq} < 0$ if and only if $\frac{p^*}{(1-\phi^*)^2} h(\widetilde{m}) < \widetilde{\alpha}$.

Part (i) of the Proposition shows that, for all distributions of the judges' and the potential violators' preferences, two of the results derived for given behavior of the counterparts carry over to the Bayesian Nash Equilibrium: The punishment frequency decreases in the magnitude of fines, and the violation frequency increases in the probability that the loss occurs even without violation.

The fact that these two results carry over to the Bayesian Nash Equilibrium is intuitive. Let us start with the first result in *part (i)* of the Proposition, $\frac{dp^*}{dF} < 0$: For all adjustments of the violators' behavior, the consequences of a type-I error are increasing in the fine size, so that judges are also, in equilibrium, more reluctant to punish in case of high fines. At the same time, however, *part (ii)* of the Proposition states that

it cannot be taken for granted that Becker's classical result concerning the deterrence effect of higher fines holds when violators anticipate the degree at which the judges' willingness to punish decreases in F. If this effect is so strong that the *expected* fine decreases, then the violation frequency increases in the fine level.

Hence, *part (ii)* of Proposition 2 confirms the results of Andreoni (1991) and Feess and Wohlschlegel (2009) that higher fines may reduce deterrence if judges care sufficiently much about type I errors. However, our result goes one step further by identifying the distribution of judges' preferences over type I and type II errors as the driving force of this result, whereas it does not depend at all on the distribution of potential violators' aversion to stealing. To see the impact of the distribution of the judges' preferences, note that a convex distribution function $G(\cdot)$ means that there is high probability mass on large α, i.e. many judges put high weight on type-I errors. And since a higher F reduces the critical threshold $\widetilde{\alpha}$, many judges do not punish for high α when $G(\cdot)$ is convex. In these cases, the indirect effect of the higher fine size via the lower punishment probability outweighs the direct effect, so that the incentive to obey the law decreases in the fine level.

We now turn to the second result in *part (i)* of the Proposition, $\frac{d\phi^*}{dq} > 0$. This means that the standard result that higher uncertainty reduces deterrence carries over from the setting with exogenously given errors in court to the Bayesian Nash Equilibrium in which judges' responses to higher uncertainty depend on their preferences. While the impact on violators is hence clear-cut, the impact on the judges' punishment frequency is ambiguous. Moreover, the impact cannot be traced back to general properties of the distribution functions, but rather depends on both equilibrium thresholds, $\widetilde{\alpha}$ and $\widetilde{m}$.

Summing up, in the Bayesian Nash Equilibrium, judges respond to the fine size and potential violators to uncertainty in the intuitive way (i.e. in the same direction as naive decision makers do), but the impact of F on the stealing frequency and of q on the punishment behavior are ambiguous.

3 Experimental design

We conducted eight sessions with a total of 192 subjects in the AIX laboratory for empirical economic studies at RWTH Aachen University. The participants consisted of 119 males and 73 females with an average age of 25 years in the range of 18 to 62 years. Sessions were conducted in September and October 2013 and were

computerized using the software z-tree (Fischbacher, 2007). On average, a session lasted approximately one hour with an average payment of 12 Euro (16US$ at the time of the experiment), including a show up fee of 4 Euro. We used ECU as the currency for the experiment with an exchange rate of 75 ECU = 0.1 Euro. A translation of the originally German instructions is provided in Appendix 2. Participants played two different roles and were informed that both parts are paid out which we did in order to avoid potentially negative (or very low) amounts for those who are innocently punished. Since we gave no feedback between the two rounds and because judges receive a fixed income anyway, there are no concerns about income effects.

The experimental design proceeds closely along the lines of the model. Participants were randomly assigned to their roles as judges (group 1) or potential violators (group 2). Then, pairs of two anonymous participants, with one judge and one potential violator each, were formed. In each role, participants got a fixed amount of 2400 ECU which was mainly done to avoid high negative payoffs for those who do not steal in their role as potential violators but who are nevertheless punished. All participants were informed that, for each member of group 2, we provide a donation of 2400 ECU to the German charity '*Brot für die Welt*', which is organized by the federation of Protestant regional church bodies in Germany, and mainly funds projects of capacity development in developing countries. The money meant for donation can be stolen by the respective member of group 2. In this case, the 2400 ECU will not be donated but are instead transferred to the account of the thief which hence increases from 2400 to 4800 ECU. The 2400 ECU that can be stolen resemble the amount A from our model. We deliberately framed the experiment as a violation of social norms by using the terms "donation", "stealing" and "fine".

All participants were informed that judges observe whether the 2400 ECU are available for donation or not, but that there is a probability of q that the money disappears even in case it is not stolen. Judges can impose exogenously given fines of F if and only if the money is not disposable for donation. In the instructions, we emphasized some features of the experiment: the meaning of q, that the money will in fact be donated, and that all features of the game are common knowledge. As this might affect their behavior, participants were not informed that they act in both roles. Of course, we controlled for order effects.

In the experiment itself, we provided the following nine combinations of q and F:

Table 1. $q - F$ matrix

q/F	1200	2400	4800
10%			
50%			
90%			

Before we started the actual experiment, we posed several control questions concerning the calculation of the payoffs. The members of both groups were then asked to make their decisions by indicating for each of the nine cells in table 1 whether they want to steal or punish, respectively. Subsequently, a second round was played with opposite roles, again by forming random pairs of judges and potential violators. Between the two rounds, no decisions or payoffs were conveyed. Finally, we distributed a personality questionnaire and a form with questions on the reasons for the decisions made. The latter form showed that participants had no problems in understanding the experiment.

4 Descriptive statistics

Judges. Since each of the 192 participants made nine decisions as judge and as violator each, we have, over all, $192 \cdot 2 \cdot 9 = 3456$ observations. Starting with judges, table 2 shows the distribution of punishments for the nine combinations of the signal's noise represented by q and fine sizes F. For each combination of q and F, we first show the absolute number of punishments. The second line shows percentages. The number in the third line is a measure of judges' incentives to punish given that they correctly anticipate thieves' actual reactions to the combinations of q and F, and is calculated as $\Delta \equiv \frac{\frac{(1-\phi)q}{\phi+(1-\phi)q}F}{\frac{\phi}{\phi+(1-\phi)q}A} = \frac{(1-\phi)Fq}{\phi A}$.[7] The numerator is the probability of a type-I error, multiplied by the fine, and hence the unjustified expected fine if a judge punishes. Similarly, the denominator is the expected amount that has been stolen unatoned in case of no punishment. Both terms are calculated for the *actual* behavior of potential violators. Δ can be interpreted

[7] For instance, given that 62.5% of violators steal in the case of $q = 50\%$ and $F = 1200$, the probability of a type-I error when punishing is $\frac{0.375 \cdot 0.5}{0.375 \cdot 0.5 + 0.625} \approx 0.23$. The expected unjustified fine in case of punishment is thus about $0.23 \cdot 1200 = 276$, whereas the expected unatoned stolen amount in case of no-punishment is about $\left(1 - \frac{0.375 \cdot 0.5}{0.375 \cdot 0.5 + 0.625}\right) \cdot 2400 \approx 1846$, the ratio of which is around 0.15.

as the ratio in the expected undesired monetary consequences with and without punishment. Recall that we defined α as the relative weight a judge puts on type-I compared to type-II errors, so that a judge who correctly anticipates Δ will punish if and only if $\alpha \leq \frac{1}{\Delta}$.

In the fourth line, we multiply Δ by $\frac{p}{1-p}$ to get the ratio of judges' expected costs of punishment and no-punishment based on the actual frequencies of each choice. Note carefully that Δ refers to the decision of a single judge, while $\Delta\frac{p}{1-p}$ is the ratio of the undesired monetary consequences of both types of error, aggregated over the actual decisions of all judges.

Table 2. Punishment behavior of judges (p)

	F=1200	F=2400	F=4800	Average
q=10%				
Frequency	81	84	84	83
Percentage	42%	44%	44%	43%
Δ	0.04	0.09	0.36	0.17
Ratio of expected costs	0.03	0.07	0.28	0.13
q=50%				
Frequency	81	87	39	69
Percentage	42%	45%	20%	36%
Δ	0.15	0.26	1.26	0.55
Ratio of expected costs	0.11	0.21	0.32	0.21
q=90%				
Frequency	56	40	47	48
Percentage	29%	21%	24%	25%
Δ	0.24	0.60	1.56	0.80
Ratio of expected costs	0.10	0.16	0.50	0.25
Average				
Frequency	73	70	57	67
Percentage	38%	37%	30%	35%
Δ	0.14	0.32	1.06	0.51
Ratio of expected costs	0.08	0.15	0.37	0.20

A first observation from table 2 is that the overall punishment frequency is rather low (35% over all nine situations) which provides preliminary evidence that judges care more about type-I errors than about type-II errors.

Turning to the impact of our primary model parameters, noise and fines, on judges' decisions, we start

with considering the impact of the fine size F. Recall from Proposition 2 that the Bayesian Nash Equilibrium predicts that the punishment frequency decreases in F. Remarkably, table 2 shows that judges only slightly respond to different fine sizes: considering the averages taken over all levels of q, the punishment frequency is 38% for $F = 1200$, 37% for $F = 2400$ and 30% for $F = 4800$. The Wilcoxon rank sum test shows that the punishment frequency for $F = 4800$ differs from the one for the other two fine levels significantly at the 5%-level, but the size effect is moderate.

Our model predicts that the punishment frequency decreases in q when judges treat the violation frequency as exogenous, (Proposition 1), but that the impact of q depends on the models' parameters in the Bayesian Nash Equilibrium (Proposition 2). Results are perfectly in line with Proposition 1: On average, the punishment frequencies are 43% for $q = 10\%$, 36% for $q = 50\%$ and 25% for $q = 90\%$, with all differences significant at the 5%-level when performing a Fisher Exact test. However, the difference between $q = 10\%$ and $q = 50\%$ is exclusively driven by the highest fine level, but insignificant for the two lower fine levels.

Now recall that the numbers in the third line are the ratios of the expected undesired monetary consequences with and without punishment, calculated as $\Delta = \frac{(1-\phi)Fq}{\phi A}$. Overall, we see that the expected monetary consequences from type-I errors are lower for seven out of the nine cases, and higher only for $F = 4800$ combined with $q = 50\%$ and $q = 90\%$. Interestingly, these are in fact two of the cases where the punishment frequencies are lowest, i.e. the observed behavior is compatible with the hypothesis that judges adjust their punishment behavior to the differences in the expected monetary consequences of the two error types.

The numbers in the fourth line show that judges indeed put far higher weight on type-I compared to type-II errors. On average, we observe $\Delta\frac{p}{1-p} = 0.20$ which means that, given the behavior of thieves *and* judges, the undesired monetary consequences of type-II errors are on average five times higher than those of type-I errors. Hence, many judges behave in line with the in dubio pro reo-principle. $\Delta\frac{p}{1-p}$ is below one in all of our nine cases, and ranges from a minimum as small as 0.03 to a maximum of 0.5.

Summing up, the descriptive statistics on judges provides two insights: First, judges respond to the exogenous variables as predicted by Proposition 1 as both higher noise (q) and higher fines (F) reduce the punishment frequency. Second, judges care far more about type-I compared to type-II errors.

Violators. Similar to table 2 for judges, table 3 displays the stealing behavior of violators. The first three lines can be interpreted analogously to those for judges: The number in the first (second) line is the frequency (percentage) of violations. The numbers in the third line are calculated as $\tau \equiv \frac{p(1-q)F}{A}$. The term in the numerator is the difference between the expected punishment, with and without violation, for the *actual* behavior of judges. The term in the denominator is simply the amount stolen. For instance, $\tau = 0.11$ for $q = 50\%$ and $F = 1200$ means that the increase in the expected fine when a participant steals amounts to only 11% of the amount stolen. Recall from the model that risk-neutral violators who anticipate the judges' decisions correctly steal whenever $1 - m \geq \tau$. In other words, everyone who is risk-neutral and puts less than $1 - 0.11 = 0.89$ weight on donation relative to his own payoff should steal.

In all nine cases, τ is below one, so that risk-neutral participants who do not care about donation should *always* steal. Given that we find, aggregated over all nine cases, $\tau = 0.22$ and a violation percentage of 0.55, it follows that many potential violators have strong social preferences or aversion against violating social norms. This impression is reinforced by the observation that the violation frequency is only around 60% for the two cases in which the fine is weakly below the amount that can be stolen.

We now consider in greater detail the impact of noise q on the violation frequency ϕ, and then turn to the fine size F. Recall that the Bayesian Nash Equilibrium predicts that ϕ increases in q. For all fine sizes, the expected rise in the punishment when a participant steals, $p\,(1-q)\,F$, decreases in q to a large degree. Thus, in line with the prediction from the model, the stealing incentive increases in the signal's noise, even when taking the actual punishment behavior of judges into account. This is confirmed for the differences between $q = 10\%$ and $q = 50\%$ and the difference between $q = 10\%$ and $q = 90\%$ (significant at the 1%-level in a Fisher exact test), but not so for the difference between $q = 90\%$ and $q = 50\%$ which is insignificant. This insignificance is driven by the fact that, for the intermediate fine size of $F = 2400$, the stealing frequency is higher for $q = 50\%$ than for $q = 90\%$.

Considering the fine size, we observe a considerable deterrence effect of the highest fine $F = 4800$ in a Fisher exact test, but no significant difference between $F = 1200$ and $F = 2400$. This holds not only on average, but also for all levels of q, that is, violators on average basically do not care whether the fine is 1200 or 2400.

Table 3. Stealing-behavior of potential violators (ϕ)

	F=1200	F=2400	F=4800	Average
q=10%				
Frequency	106	101	68	92
Percentage	55%	53%	35%	48%
τ	0.19	0.39	0.79	0.46
q=50%				
Frequency	120	127	85	111
Percentage	63%	66%	44%	58%
τ	0.11	0.23	0.20	0.18
q=90%				
Frequency	125	115	103	114
Percentage	65%	60%	54%	60%
τ	0.01	0.02	0.05	0.03
Average				
Frequency	117	114	85	106
Percentage	61%	60%	44%	55%
τ	0.10	0.21	0.35	0.22

Summing up, the descriptive statistics suggests that, on average, the participants respond to noise q and fine size F in the directions predicted by our model. In the subsequent sections, we use more rigorous methods to analyze the participants' responses to q and F. We will focus on two issues:

First, we are interested in the interaction of q and F, that is, how the impact of higher fines on the participants' behavior is moderated by noise. To see the point, recall from table 2, that for instance the lower punishment frequency for $F = 4800$ compared to $F = 2400$ holds only for $q = 50\%$, but not so for the other two levels of noise. The interaction of q and F can best be analyzed in regressions of the punishment and stealing behavior, and this will be done in section 5.

Second, a closer look at our raw data reveals that there is a large heterogeneity in the behavior of participants. We address this issue in two ways: On the one hand, we include personal characteristics derived from a questionnaire (see below) in our regression analysis, to see whether these characteristics can partially explain the observed behavior. On the other hand, we perform a cluster analysis to distinguish between different types of participants (section 6).

5 Regression analysis

Factor analysis on personal characteristics. To control for potentially relevant personal characteristics, we asked participants to complete a questionnaire after the experiment. The questions referred to risk attitudes, moral attitudes towards the violation of legal and social norms and to the willingness to punish. These questions were tailored specifically to this experiment and are listed in Appendix 3. We performed a confirmatory factor analysis (see e.g. Jae-on and Mueller, 1978) which led to four factors used in our regression analyses:

- The first factor which we refer to as "attitude to risk" comprises three questions mainly regarding financial risk attitudes (willingness to invest into a mutual fund, gambling in a poker game, and investing into a startup).
- The second factor which we denote as "moral attitudes" consists of three questions regarding honest behavior (minor wrong statements in tax declaration, plagiarism, keeping a found purse with 200 Euro).
- Factors three and four are determined from a questionnaire on the determinants of why people follow rules. Factor three comprises two questions measuring the impact of fines and consequences, and factor four refers more generally to the degree to which people are self-responsible for their actions. A detailed list of questions is provided in Appendix 3.

Regression analysis for judges. In all regressions, we control for order effects as it might influence the behavior whether the role as judge or as a potential thief is played first. Reference category is "judge first", and reference category for gender is "male". We also include the personality factors just described in all regressions. In the regressions in columns (1), (2) and (4), the reference category for q and F are the intermediate values, that is, $q = 50\%$ and $F = 2400$. All coefficients are marginal effects, evaluated at the mean of the explanatory variable. Furthermore, we need to account for the fact that each participant makes nine decisions as a judge. Since these nine decisions are correlated, we cluster our observations in all regression on a subject level.

Table 4 shows results for a probit-model on the behavior of judges.[8] In the first two columns, we regress

[8] All results are qualitatively the same for logit-models.

Table 4. Probit-model on behavior of judges

	(1)	(2)	(3)
Age	-0.007**	-0.007**	-0.007**
	(0.003)	(0.003)	(0.003)
Gender = Female	0.057	0.058	0.057
	(0.040)	(0.041)	(0.040)
Δ			-0.168***
			(0.028)
$F = 1200$	0.014	-0.029	
	(0.020)	(0.039)	
$F = 4800$	-0.076***	-0.252***	
	(0.019)	(0.034)	
$q = 10\%$	0.075***	-0.015	
	(0.028)	(0.044)	
$q = 90\%$	-0.118***	-0.243***	
	(0.027)	(0.038)	
$F_{1200} \cdot q_{10\%}$		0.015	
		(0.053)	
$F_{1200} \cdot q_{90\%}$		0.134**	
		(0.066)	
$F_{4800} \cdot q_{10\%}$		0.290***	
		(0.062)	
$F_{4800} \cdot q_{90\%}$		0.333***	
		(0.062)	
More than 3 thefts	0.134***	0.135***	0.134***
	(0.039)	(0.040)	(0.039)
Attitude to risk	-0.029	-0.030	-0.029
	(0.021)	(0.021)	(0.021)
Moral attitudes	0.059***	0.059***	0.059***
	(0.022)	(0.022)	(0.022)
Fines and consequences	0.032*	0.032*	0.032*
	(0.019)	(0.019)	(0.019)
Self responsibility	0.054***	0.054**	0.054***
	(0.021)	(0.021)	(0.021)
Order = Thief first	0.035	0.036	0.035
	(0.038)	(0.039)	(0.038)
Observations	1728	1728	1728

One (two, three) stars denote significance at the ten- (five, one-) percent level, respectively. All coefficient marginal effects. Standard error in brackets.

the punishment probability only on variables which are not affected by the actual behavior of violators. In line with the descriptive statistics, we find that judges do not differentiate between the low and the medium fine size (the latter one is the reference category), but an increase from 2400 to 4800 reduces the punishment probability by 7.5 percentage points. Given that the average punishment probability in the data is 35%, this amounts to a notable reduction of about 21 percent. For q, we also adopt the intermediate value as reference category, and our results confirm that judges reduce their punishment frequencies significantly when the probability that the money disappears, even when it is not taken by potential violators, increases. Overall, judges respond in the direction predicted by Proposition 1.

When we add the interaction terms of q and F in column (2), we find no difference between $q = 10\%$ and the reference category of $q = 50\%$ for the lowest fine, but all other interaction terms are significant. In particular, the negative impact of the signal's noise on the punishment frequency is most pronounced for the intermediate fine level of $F = 2400$,

In column (3), the variable on the right hand side of the regression that we are mainly interested in is $\Delta \equiv \frac{(1-\phi)Fq}{\phi A}$, which is the ratio in the expected unjustified fine from type-I errors in case of punishment to the unatoned theft from type-II errors in case of no-punishment, both calculated using potential thieves' actual decisions. That is to say, the fine size F and the noise q, which are not explicitly considered as controls in column 3, and the actual stealing frequency ϕ are the determinants of Δ. We find that judges respond significantly to this difference.

Some interesting results emerge for our additional controls. First, we have a dummy variable which indicates whether a judge has, in his role as potential violator, stolen less or weakly more than three times. We had no prior on the sign of this dummy: On the one hand, one might assume that those who steal more often find theft more acceptable, and hence punish less often. On the other hand, they might assume a higher frequency of theft, which hence reduces the level of Δ that they expect. Controlling for other factors such as moral attitudes, we find that those who steal more often also punish more often. Furthermore, in line with the intuition, those with higher moral attitudes and those who believe more in social responsibility punish more often (recall the details for these factors described above). Older students punish less often, and gender is insignificant.

Summing up our findings for judges, our first and most fundamental result is that the punishment probability responds strongly to the signal's noise and the fine size. Note that this implies that the ratio of type-I and type-II errors cannot be treated as independent of these parameters, which is usually assumed in the literature on errors in court discussed in the introduction. Second, the effects of q and F are both in line with our model predictions where the impact of the fine size, however, is exclusively driven by the highest fine. Third, the violation frequency in fact depends on the ratio of the undesired consequences of type-I and type-II errors, that is, it is decreasing in Δ. Fourth, we find that those personal characteristics that are most closely related to our setting are highly significant as participants who believe in moral and self responsibility punish more often.

Regression analysis for thieves. Turning to thieves, we again start by regressing the behavior on the fine size F and the signal's noise captured by q, that is, we do not take the punishment frequency of judges explicitly into account. For both q and F, the intermediate values serve as reference categories. The model's prediction that the violation frequency increases in q is only confirmed for the comparison of the two lower values of q, but there is no significant difference between $q = 50\%$ and $q = 90\%$. As judges do, violators do not differentiate their behavior between fines of 1200 and 2400, but increasing the fine from 2400 to 4800 reduces the violation frequently sharply, by about 15 percentage points or 27 percent.

When we add the interaction terms of q and F in column (2), we find no difference between $q = 10\%$ and the reference category of $q = 50\%$, but both interaction terms with $q = 90\%$ are positive. Thus, the impact of increasing the signal's noise from $q = 50\%$ to $q = 90\%$ on the stealing frequency is lowest for the intermediate fine level of $F = 2400$, which matches the corresponding result for judges.

In column (3), we substitute q and F by $\tau = \frac{p(1-q)F}{A}$, the ratio of the expected punishment with and without violation for the actual behavior of judges in the numerator, and the amount stolen. Participants respond strongly to incentives: an increase in the ratio of expected punishment with and without violation reduces the stealing frequency significantly.

Summing up, we find that violators respond to the signal's noise and the fine size in directions as predicted by our model. These results, however, are driven by the difference between $q = 10\%$ and the two other levels of uncertainty, and between $F = 4800$ and the two lower fine levels, respectively. Thus, only a particularly

Table 5. Probit-model on behavior of thieves

	(1)	(2)	(3)
Age	-0.006	-0.006	-0.006
	(0.005)	(0.005)	(0.005)
Gender = Female	-0.119**	-0.119**	-0.117**
	(0.050)	(0.050)	(0.050)
τ			-0.330***
			(0.062)
$F = 1200$	0.016	-0.040	
	(0.023)	(0.040)	
$F = 4800$	-0.160***	-0.231***	
	(0.025)	(0.039)	
$q = 10\%$	-0.105***	-0.145***	
	(0.028)	(0.043)	
$q = 90\%$	0.021	-0.068*	
	(0.028)	(0.040)	
$F_{1200} \cdot q_{10\%}$		0.066	
		(0.042)	
$F_{1200} \cdot q_{90\%}$		0.097**	
		(0.045)	
$F_{4800} \cdot q_{10\%}$		0.049	
		(0.046)	
$F_{4800} \cdot q_{90\%}$		0.158***	
		(0.041)	
More than 3 convictions	0.110**	0.110**	0.109**
	(0.052)	(0.052)	(0.052)
Attitude to risk	-0.011	-0.011	-0.011
	(0.025)	(0.025)	(0.024)
Moral attitudes	0.018	0.019	0.018
	(0.026)	(0.026)	(0.026)
Fines and consequences	0.046*	0.046*	0.046*
	(0.026)	(0.026)	(0.026)
Self responsibility	-0.066***	-0.066***	-0.065***
	(0.025)	(0.025)	(0.025)
Order = Thief first	-0.048	-0.048	-0.047
	(0.048)	(0.049)	(0.048)
Observations	1728	1728	1728

One (two, three) stars denote significance at the ten- (five, one-) percent level, respectively. All coefficient marginal effects. Standard error in brackets.

Table 6. Punishment and violation frequencies; disaggregated by individuals

Number of thefts or punishments per individual	Frequencies (judges)	Frequencies (violators)
0	20%	9%
1	9%	5%
2	10%	9%
3	26%	**16%**
4	6%	4%
5	11%	11%
6	**11%**	14%
7	2%	8%
8	1%	6%
9	4%	19%

low uncertainty and a particularly high fine deter participants from violation. Interestingly, self responsibility is significant in the expected direction, but moral attitudes are not. Similar to the regression on judges, we find that those who punish more often also violate more often. The stealing probability of females is by 11.9 percentage points or 20 percent lower than those of males, significant at the 5%-level.

6 Heterogeneity of the participants

So far, we have restricted attention to the average behavior of judges and potential violators, but we have not yet considered the individual behavior. Individual behavior is important for several reasons: first, participants may largely differ in the relative weight they put on type-I and type-II errors. Second, some participants may behave more or less independently of q and F by stealing (or punishing) basically always or never. Then, the low differences in the violation frequency for different levels of q can potentially be driven by the fact that many individuals do not adjust their behavior at all. Table 6 summarizes the individual behavior of the participants.

The first column captures the number of punishments and violations, respectively, per individual. The second and the third column measure the number of individuals meeting these frequencies. For instance, the bold "11%" expresses that 11% of all judges penalized exactly six times, and the bold "16%" means that 16% of all participants steal exactly three times. The table shows that 24% percent of all judges (20% of punish never and 4% always) and 28% of all potential violators (9% steal never and 19% always) behave

identically for all levels of noise and fines in our experiment.

To learn more about the differences in the behavior of subgroups, we perform a cluster analysis. Based on the nine decisions in their roles as judges and thieves, respectively, we adopted an average linkage cluster approach which generates clusters based on the observations' average distance to each other. Observations with the smallest average distance form clusters (see e.g. Everitt et al., 2011). Participants who punish and steal either always or never form separate clusters.

Judges. The largest Cluster 1 (38.5%) consists of judges whose punishment frequencies are intermediate and almost independent of the fine and the noise of the signal. They seem to deviate from this pattern only where fines and noise are both high, in which case expected preference costs caused by type-I errors are highest compared to those of type-II errors (see table 2 above) and, therefore, judges punish very rarely. However, there is one observation that can't be explained by this argument: For the very highest levels of fine and noise, judges in this cluster return to the intermediate frequency with which they punish for low fines and noise.

The second largest cluster (24.5%) consists of judges who punish basically always for low q, frequently for medium q, and never for high q. By contrast, their behavior is more or less independent of F. The large impact of q indicates that these judges take q as a very good predictor for the stealing probability, that is, they put low emphasis on the possibility that violators adjust their behavior to q.

The third cluster consists of the 20.3 percent of judges who never punish.

All other clusters are fairly small: Cluster 4 consists of judges whose punishment frequency is to a large degree increasing in q, a behavior that is difficult to rationalize. The same holds for the 4.7% of judges in cluster 5 who punish far most frequently for the intermediate fine level of $F = 2400$. Finally 4.2% of judges punish whenever they observe that the money is gone. Recalling from table 2 that Δ_I^{II} is negative in all but two cases, such a behavior can be rationalized when the weight put on type-II errors is (slightly) higher compared to type-I errors.

Thieves The cluster analysis for thieves also leads to six different patterns. The largest Cluster 1 consists of thieves whose behavior is in line with theoretical predictions: The higher the noise and the lower the fine,

Table 7. Cluster analysis for judges

		F=1200	*F=2400*	*F=4800*
Cluster 1	*q=10%*	28%	30%	34%
N=74	*q=50%*	31%	36%	5%
(38.5%)	*q=90%*	35%	12%	30%
Cluster 2	*q=10%*	100%	98%	98%
N=47	*q=50%*	68%	62%	30%
(24.5%)	*q=90%*	15%	2%	4%
Cluster 3	*q=10%*	0%	0%	0%
N=39	*q=50%*	0%	0%	0%
(20.3%)	*q=90%*	0%	0%	0%
Cluster 4	*q=10%*	33%	27%	27%
N=15	*q=50%*	87%	100%	60%
(7.8%)	*q=90%*	100%	100%	80%
Cluster 5	*q=10%*	0%	44%	11%
N=9	*q=50%*	56%	89%	44%
(4.7%)	*q=90%*	0%	78%	33%
Cluster 6	*q=10%*	100%	100%	100%
N=8	*q=50%*	100%	100%	100%
(4.2%)	*q=90%*	100%	100%	100%

the higher is the stealing frequency (with the exemption at the intermediate values of q and F where the stealing frequency is 100%). Given the actual behavior of judges, this is also consistent with the regression analysis which shows that the violation frequency decreases in the difference between the expected fine with and without violation (see table 5, column I).

The second largest cluster consists of violators whose behavior depends mainly on the signal's noise, in the expected direction: The higher q, the higher the violation frequency. The difference to the first cluster is that higher fines still reduce the violation frequency, but are far less important. In both of these largest

clusters, however, violators respond in the expected directions on noise and fine sizes.

18.8% of the participants steal always (cluster 3).

The remaining three clusters are small: Cluster 4 consists of participants whose behavior can hardly be rationalized, since the violation frequency is to a large extent decreasing in q, and because the fine size plays basically no role. 8.9% of all participants never steal (cluster 5), and the behavior in cluster 6 is almost completely driven by the fine size: These eleven participants (5.7%) almost always steal for low F, but hardly ever for the two higher fine sizes.

Table 8. Cluster analysis for thieves

		F=1200	*F=2400*	*F=4800*
Cluster 1	*q=10%*	72%	52%	17%
N=58	*q=50%*	95%	100%	40%
(30.2%)	*q=90%*	100%	84%	66%
Cluster 2	*q=10%*	2%	22%	8%
N=50	*q=50%*	16%	46%	34%
(26.0%)	*q=90%*	48%	56%	48%
Cluster 3	*q=10%*	100%	100%	100%
N=36	*q=50%*	100%	100%	100%
(18.8%)	*q=90%*	100%	100%	100%
Cluster 4	*q=10%*	100%	100%	90%
N=20	*q=50%*	50%	50%	45%
(10.4%)	*q=90%*	10%	10%	20%
Cluster 5	*q=10%*	0%	0%	0%
N=17	*q=50%*	0%	0%	0%
(8.9%)	*q=90%*	0%	0%	0%
Cluster 6	*q=10%*	64%	36%	0%
N=11	*q=50%*	100%	0%	0%
(5.7%)	*q=90%*	45%	0%	9%

7 Conclusion

We analyze the impact of fine size and legal uncertainty on the frequencies of punishments and legal infringements. In our theoretical model, we assume that legal decision makers have heterogeneous preferences with respect to type-I and type-II errors, and that potential violators have different preference costs from violating the law. Based on these assumptions, we first derive the straightforward results that higher uncertainty and higher fines reduce the punishment frequency when legal decision makers ignore the strategic interdependency between their own and the violators' decisions (non-strategic behavior). Analogously, higher uncertainty increases and higher fines reduce the violation frequency. Taking the interdependency between the two market sides seriously, however, results turn out to be more complicated: higher fines still reduce the punishment frequency and higher uncertainty increases the violation frequency, but the impact of the fine size on the violation frequency and the impact of uncertainty on the punishment frequency is more involved and depends on the parameters of the model.

We then conduct a laboratory experiment that accounts for differences in uncertainty and fine sizes. Our findings are basically in line with the theory, and the following results are most notable: First, in their role as judges, participants care far more about type-I compared to type-II errors which leads to rather low punishment frequencies: The undesired monetary consequences of type-II errors are five times higher than those of type-I errors. This means that the average preferences of the students participating in our experiment coincide with the principle of in dubio pro reo that is anchored in all legal systems. Second, although the expected own monetary payoff is always higher with stealing, the average stealing probability is only 55%. We hence find pronounced social preferences for donation or for compliance with social norms. Third, the data reveals a large heterogeneity in preferences both in the role as a judge and in the role as a potential violator.

Compared to the literature which treats the (relative) frequencies of type-I and type-II errors as exogenously given, two of our findings deserve attention from an applied point of view: First, when the signal's noise represented by q in our setting, increases, then there are two detrimental effects on deterrence: The first effect is that both error types reduce the difference in the expected fine with and without infringement, and this sets higher violation incentives as analyzed in the traditional literature. In addition, however, judges in

our experiment are less willing to punish in cases of higher legal uncertainty, and this indirect effect reinforces the negative deterrence effect. Second, although the results indicate that the deterrence effect of higher fines is likely to be overestimated when the countervailing effect via the lower willingness to convict a suspect is neglected; precisely as emphasized by the old dignified legal literature on nullification mentioned in the introduction.

References

[1] Ambrus, A., & Greiner, B. (2012). Imperfect Public Monitoring with Costly Punishment: An Experimental Study. *The American Economic Review, 102*(7), 3317-3332.

[2] Andreoni, J. (1991). Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty fit the Crime? *The RAND Journal of Economics, 22*(3), 385-395.

[3] Becker, G. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*(2), 169-217.

[4] Bornstein, G., & Weisel, O. (2010). Punishment, Cooperation, and Cheater Detection in "Noisy" Social Exchange. *Games, 1*(1), 18-33.

[5] Casari, M., & Luini, L. (2009). Cooperation under Alternative Punishment Institutions: An Experiment. *The Journal of Economic Behavior & Organization, 71*(2), 273-282.

[6] DeAngelo, G., & Charness, G. (2012). Deterrence, Expected Cost, Uncertainty and Voting: Experimental Evidence. *Journal of Risk and Uncertainty, 44*(1), 73-100.

[7] Egas, M., & Riedl, A. (2008). The Economics of Altruistic Punishment and the Maintenance of Cooperation. *Proceedings of the Royal Society B: Biological Sciences, 275*(1637), 871-878.

[8] Ertan, A., Page, T. & Putterman, L. (2009). Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem. *European Economic Review, 53*(5), 495-511.

[9] Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th Edition). London: John Wiley & Sons.

[10] Falk, A., & Fischbacher, U. (2002). "Crime" in the Lab-detecting Social Interaction. *European Economic Review, 46*(4), 859-869.

[11] Feess, E., & Wohlschlegel, A. (2009). Why Higher Punishment May Reduce Deterrence. *Economic Letters, 104*(2), 69-71.

[12] Fehr, E., & Fischbacher, U. (2004). Third-party Punishment and Social Norms. *Evolution and Human Behavior, 25*(2), 63-87.

[13] Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review, 90*(4), 980-994.

[14] Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics, 10*(2), 171-178.

[15] Grechenig, K., Nicklisch, A., & Thöni, C. (2010). Punishment Despite Reasonable Doubt: A Public Goods Experiment with Sanctions under Uncertainty. *Journal of Empirical Legal Studies, 7*(4), 847-867.

[16] Harrison, G., & Rutström, E. (2008). Risk Aversion in the Laboratory. *Research in Experimental Economics, 12*, 41-196.

[17] Hylton, K., & Khanna, V. (2007). A Public Choice Theory of Criminal Procedure. *Supreme Court Economic Review, 15*(1), 61-118.

[18] Jae-on, K., & Mueller, C. (1978). *Factor Analysis. Statistical Methods and Practical Issues.* Sage Publications.

[19] Khadjavi, M. (2014). On the Interaction of Deterrence and Emotions. *Journal of Law, Economics and Organization (forthcoming).*

[20] Lando, H. (2006). Does Wrongful Conviction lower Deterrence? *The Journal of Legal Studies, 35*(2), 327-338.

[21] Masclet, D., Noussair, C., Tucker, S. & Villeval, M.C. (2003). Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism. *The American Economic Review, 93*(1), 366-380.

[22] Miceli, T. (1991). Optimal Criminal Procedure: Fairness and Deterrence. *International Review of Law and Economics, 11*(1), 3-10.

[23] Michael, J., Wechsler, H. (1937). A Rationale of the Law of Homicide II. *Columbia Law Review 37*(8), 1261-1325.

[24] Nikiforakis, N., & Normann, H. (2008). A Comparative Statics Analysis of Punishment in Public-good Experiments. *Experimental Economics, 11*(4), 358-369.

[25] Nikiforakis, N. (2008). Punishment and Counter-punishment in Public Good Games: Can We Really Govern Ourselves? *Journal of Public Economics, 92*(1), 91-112.

[26] Persson, M., & Siven, C. (2007). The Becker Paradox and Type I versus Type II Errors in the Economics of Crime. *International Economic Review, 48*(1), 211-233.

[27] Png, I. (1986). Optimal Subsidies and Damages in the Presence of Judicial Error. *International Review of Law and Economics, 6*(1), 101-105.

[28] Polinsky, A., & Shavell, S. (1999). The Economic Theory of Public Enforcement of Law. *Journal of Economic Literature, 38*(1), 45-76.

[29] Rega, M., & Telle, K. (2004). The Impact of Social Approval and Framing on Cooperation in Public Good Situations. *Journal of Public Economics, 88*(7), 1625-1644.

[30] Rizzolli, M., & Saraceno, M. (2013). Better That Ten Guilty Persons Escape: Punishment Costs Explain the Standard of Evidence. *Public Choice, 155*(3-4), 395-411.

[31] Rizzolli, M., & Stanca, L. (2012). Judicial Errors and Crime Deterrence: Theory and Experimental Evidence. *Journal of Law and Economics, 55*(2), 311-338.

[32] Schildberg-Hörisch, H., & Strassmair, C. (2012). An Experimental Test of the Deterrence Hypothesis. *Journal of Law, Economics, and Organization, 28*(3), 447-459

[33] Volokh, A. (1997). n Guilty Men. *University of Pennsylvania Law Review, 146*(2), 173-216.

Appendix 1: Proof of Proposition 2

If S_α and S_m are intervals, $\widetilde{\alpha} \in S_\alpha$ and $\widetilde{m} \in S_m$, then the Bayesian Nash equilibrium is given by the solution to the system of equations

$$\begin{aligned} p &= G\left(\frac{\phi}{q(1-\phi)}\frac{A}{F}\right) \\ \phi &= H\left(1-\frac{Fp(1-q)}{A}\right). \end{aligned}$$

The system of total differentials is

$$\begin{pmatrix} 1 & -g(\widetilde{\alpha})\frac{A}{(1-\phi)^2 qF} & g(\widetilde{\alpha})\frac{\phi A}{(1-\phi)qF^2} & g(\widetilde{\alpha})\frac{\phi A}{(1-\phi)q^2 F} \\ h(\widetilde{m})\frac{F(1-q)}{A} & 1 & h(\widetilde{m})\frac{p(1-q)}{A} & -h(\widetilde{m})\frac{Fp}{A} \\ & & & \end{pmatrix} \begin{pmatrix} dp \\ d\phi \\ dF \\ dq \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{7}$$

Hence, the comparative statics are:

$$\frac{d\phi}{dF} = \frac{h(\widetilde{m})(1-q)\left(g(\widetilde{\alpha})\widetilde{\alpha}-p\right)}{A\left(1+\frac{1-q}{q(1-\phi)^2}h(\widetilde{m})g(\widetilde{\alpha})\right)}, \tag{8}$$

the denominator of which is always positive, so that the whole expression is positive if and only if $p < \frac{A\phi}{Fq(1-\phi)}g(\widetilde{\alpha})$, which is, in equilibrium, equivalent to $G(\widetilde{\alpha}) < \widetilde{\alpha}g(\widetilde{\alpha})$, which is always satisfied if $G(.)$ is convex.

$$\frac{dp}{dF} = -\frac{\widetilde{\alpha}g(\widetilde{\alpha})}{F}\cdot\frac{1+\frac{Fp(1-q)}{\phi(1-\phi)A}h(\widetilde{m})}{1+\frac{\widetilde{\alpha}F(1-q)}{\phi(1-\phi)A}g(\widetilde{\alpha})h(\widetilde{m})} < 0 \tag{9}$$

independent of the distributions of α and m.

$$\frac{d\phi}{dq} = \frac{\frac{\widetilde{\alpha}}{q}g(\widetilde{\alpha})+\frac{p}{1-q}}{\frac{\widetilde{\alpha}}{\phi(1-\phi)}g(\widetilde{\alpha})+\frac{A}{F(1-q)h(\widetilde{m})}} > 0 \tag{10}$$

independent of the distributions of α and m.

$$\frac{dp}{dq} = -\frac{\frac{g(\widetilde{\alpha})}{q}\left[\widetilde{\alpha} - \frac{p}{(1-\phi)^2}h(\widetilde{m})\right]}{1 + g(\widetilde{\alpha})h(\widetilde{m})\frac{1-q}{q(1-\phi)^2}}, \tag{11}$$

the denominator of which is always positive, so that we have $\frac{dp}{dq} < 0$ if and only if

$$\frac{p^*}{(1-\phi^*)^2}h(\widetilde{m}) < \widetilde{\alpha}. \tag{12}$$

■

Appendix 2: Translation of German instructions

Welcome to this experiment and thank you very much for your participation! This experiment has been financed by researchers from RWTH Aachen, Frankfurt School of Finance & Management and Portsmouth Business School.

Please turn off your mobile phones and remain silent during the entire experiment. Any communication between you and the other participants is not allowed. If you have questions, please raise your hand. We will then directly come to your cabin to answer your questions.

The instructions are written using the masculine form only in order to improve readability. Please understand this as being gender-neutral.

All of your decisions will be processed anonymously and cannot be traced back to you.

During the experiment all amounts will be presented in ECU (Experimental Currency Unit). At the end of the experiment the amount of ECU will be converted to Euro as follows:

75 ECU = 10 Cent (EUR)

You will receive a show up fee of 3000 ECU for participating in this experiment.

The experiment consists of two rounds. Your final payment will be the sum of your payments from all two rounds and your show up fee.

During the experiment you are allowed to use any tools such as paper, pencils or calculators.

After the actual experiment we will ask you to fill out a questionnaire. Please answer these questions honestly. The answers to the questionnaire will not impact your payment.

All participants have been randomly assigned to one out of two groups. You have been assigned to Group 1. In your role you will have to take decisions, which do not have any impact on your payment in the first round. Your payment in this round will be 2400 ECU independently of your decisions. Nevertheless we ask you to take this round seriously.

In round 1 the computer randomly assigns one participant from Group 2 to you. This participant will also receive 6000 ECU for participating in this experiment and 2400 ECU in round 1.

The experiment is as follows: For every participant in Group 2 (including the person that has been assigned to you) we have provided a donation of another 2400 ECU to "Brot für die Welt". However, the participants in Group 2 have the possibility to steal this donation. In this case, the 2400 ECU will not be donated to "Brot für die Welt". At the same time the payment for the participant in Group 2 will be increased from 2400 ECU to 4800 ECU.

You will be able to see whether the donation of 2400 ECU is available or not. The only complication is that you cannot be entirely sure whether the money has been stolen by the participant in Group 2: After the potential thief has decided about stealing the donation, the 2400 ECU can also get lost by chance. The probability that the donation gets lost by chance will be varied but always be public knowledge.

This means concretely: If you notice, that the 2400 ECU are not available for donation, you cannot be sure whether the donation has been stolen or been lost by chance.

In case that the donation is not available (and only in this case) you can decide to punish the participant from Group 2 that has been assigned to you. However, in this case you have to consider that you might punish an innocent person. If you decide not to punish a potential thief might not get

any a punishment for stealing the donation.

Please consider: In case that the money is still available we will really donate the money!

The punishments will be varied as well but also always be public knowledge.

You and the person that has been assigned to you will see the following table during the experiment:

q/F	1200	2400	4800
10%			
50%			
90%			

Following the lines from the top you see the probability q. This is the probability that the donation gets lost by chance, even if the money has not been stolen.

Example: In the second line the 2400 ECU will not be available for donation with a 50% chance even if the person from Group 2 has not stolen the donation.

Let us clarify this: Considering that the donation is not available, the probability that an innocent person gets punished will be higher in a lower line (if you decide to punish).

Following the columns from the left you can see the different amounts of punishment. A punishment of 4800 ECU means that 4800 ECU will be subtracted from the account of the other participant. This punishment will only become relevant if the money is not available. For this case we ask you to decide in which cases of q (probability that donation gets lost by chance if it has not been stolen) and amount of punishment you want to punish.

Before you fill out this table, please answer the following question:

Suppose you consider the probability that the person assigned to you steals with 20% probability. You know that the donation gets lost with a 50% chance (q=50%) even if the money has not been stolen. You then notice that the donation is not available.

How would you estimate the actual probability that the money got stolen? If you need a calculator please use the icon on the right. Please type in a value, even if you are unsure about your result.

Let us quickly sum up:

Assume that the donation of 2400 ECU is not available. Following the lines from the top you can see the probability q that the donation got lost by chance (if it has not been stolen). In the columns you can see possible punishments.

The participant that has been assigned to you has to decide whether he wants to steal or not for every case. If he has stolen and does not get punished and receives a higher payment. If he has not stolen but the donation has been lost by chance, you might punish an innocent person.

Round 1: Judge Please mark the cases in which you want to punish. You have to decide for every case individually. A checkmark means that you want to punish, a blank field means that you do not want to punish.

Round 2: Thief In the following round you take the role of the participants in Group 2. Apart from this, there are no changes in the experiment compared to round 1. You receive a base payment of 2400 ECU in this round. A checkmark in the table now means that you steal the donation

of 2400 ECU. If the random participant from the other group decides to punish you, the punishment will be subtracted from your account.

Before the experiment begins we would like to explain how the actual payment in round 2 is calculated. We ask you again to fill the table with all nine combinations of q (probability that the donation gets lost, even if you decide not to steal) and the amount of punishment. At the same time we will ask a person from the other group to decide in which of the nine cases he wants to punish if the donation is not available.

For the actual payment the computer randomly selects one of the nine cases; each case with the same probability.

Example: Assume the case with q=50% and a punishment of 2400 ECU is selected. If the donation is still available, your payment will be 2400 ECU. If the donation is not available, there are four possibilities:

(1) You have stolen and you get punished. Your payment then will be: 2400 ECU (base payment) + 2400 ECU (stolen donation) - 2400 ECU (punishment) = 2400 ECU

(2) You have not stolen but you get punished. Your payment then will be: 2400 ECU (base payment) + 0 ECU (donation not stolen) - 2400 ECU (punishment) = 0 ECU

(3) You have stolen and you do not get punished. Your payment then will be: 2400 ECU (base payment) + 2400 ECU (stolen donation) = 4800 ECU

(4) You have not stolen and you do not get punished. Your payment then will be: 2400 ECU (base payment) + 0 ECU (donation not stolen) = 2400 ECU

Please mark the cases in which you want to steal. You have to decide for every case individually. A checkmark means that you want to steal, a blank field means that you do not want to steal.

Appendix 3: Questionnaire

Attitude to risk:

How probably would you decide to...

... invest 10 % of your yearly income in an open mutual fund with medium growth opportunities?

... invest your daily income in a poker game?

... invest 10 % of your yearly income into an entrepreneurial company?

Moral attitudes:

How probably would you decide to...

... state favorable yet questionable information in your tax declaration?

... declare someone else's work as your own?

... keep a found purse with 200 Euro?

Impact of fines and consequences:

To which degree do you agree to the following statements?

Whether people follow rules, depends mainly on the consequences

For many crimes punishment in Germany is too low.

Necessity of rules and regulation:

To which degree do you agree to the following statements?

The financial crisis has been caused because risks were not disclosed and underestimated.

People should be held responsible for their actions.

Self-responsibility for own actions:

To which degree do you agree to the following statements?

Whether people follow rules, depends mainly on their character and their general living conditions.

Everyone is responsible for him-/herself.

Donations are a relevant component to fight poverty and misery.

Fault of financial crisis:

To which degree do you agree to the following statements?

The financial crisis has been caused because bankers have taken risks on behalf of the community to enrich themselves.

The financial crisis has been caused because of bad regulation.