JOURNAL OF
EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 7, Issue 4, 847–867, December 2010

Punishment Despite Reasonable Doubt—A Public Goods Experiment with Sanctions Under Uncertainty

*Kristoffel Grechenig, Andreas Nicklisch, and Christian Thöni**

Under a great variety of legally relevant circumstances, people have to decide whether or not to cooperate when they face an incentive to defect. The law sometimes provides people with sanctioning mechanisms to enforce pro-social behavior. Experimental evidence on voluntary public goods provision shows that the option to punish others substantially improves cooperation, even if punishment is costly. However, these studies focus on situations where there is no uncertainty about the behavior of others. We investigate sanctions in a world with "reasonable doubt" about the contributions of others. Interestingly, people reveal a high willingness to punish even if their information about cooperation rates is highly inaccurate. If there is some nontrivial degree of noise, punishment (1) cannot establish cooperation high and (2) reduces welfare even below the level of a setting without punishment. Our findings suggest that sufficient information accuracy about others' behavior is crucial for the efficiency of sanction mechanisms. If a situation is characterized by low information accuracy, precluding sanctions, for example, through high standards of proof, is likely to be optimal.

I. Introduction

Far be it from you to do such a thing—to kill the righteous with the wicked, treating the righteous and the wicked alike. . . . The Lord said, "If I find fifty righteous people in the city of Sodom, I will spare the whole place for their sake."

Genesis 18, 25–26

Sanctions are a key element of justice, where interventions try to reduce incentives for misbehavior. An ideal setting would provide perfect information, that is, people who have

*Address correspondence to Kristoffel Grechenig, Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Strasse 10, D-53113 Bonn, Germany; email: grechenig@coll.mpg.de. Grechenig is Senior Research Fellow at the Max Planck Institute for Research on Collective Goods in Bonn, Germany; Nicklisch is Assistant Professor at the University of Hambury, Germany; Thöni is Assistant Professor at the University of St. Gallen, Switzerland.

The authors thank Brian Cooper, Christoph Engel, Sven Fischer, Bruno Frey, Sally Gschwend, Georg von Heusinger, Matthias Lang, Brad LeVeck, Mathew McCubbins, Matteo Rizzolli, Kathy Spier, an anonymous referee, the participants of the 4th Annual Conference on Empirical Legal Studies at USC, the participants of the Ratio Discussion Group at the Max Planck Institute for Research on Collective Goods, the participants of the Frey-Frey-Engel-Workshop, the participants of the Economics Science Association Meeting in Copenhagen, and the participants of the 20th Annual Meeting of the American Law and Economics Association at Princeton.

to decide whether to impose sanctions would be aware of all relevant facts. Reality is much less perfect. Decisionmakers typically face imprecise, contradictory, or even wrong information. This raises the issue of whether the premise that people are imperfectly or noisily informed systematically influences their disposition to use sanctions. Using an experimental approach, we contribute to this debate by studying the effects of a ceteris paribus variation of the accuracy of information about other people's behavior. First, we ask how the fact that people receive noisy information about the behavior of those to be sanctioned affects their willingness to execute sanctions (punishment). Second, we examine how sanctioned persons respond to punishment that they receive under a noisy information system. Third, we analyze how sanctions affect cooperation and efficiency in this situation.

The implications of imperfect or noisy information on sanctioning mechanisms have received some attention in the literature.[1] Lawyers have often made intuitive statements, most famously William Blackstone (1765–1769:bk 4, ch. 27), indicating that "the law holds that it is better that ten guilty persons escape than that one innocent suffer." Others have offered both larger and smaller ratios, including Justice Cardozo (1916) (in *People v. Galbo* with reference to Lord Hale), saying: "it is better five guilty persons should escape unpunished, than one innocent person should die"), and Benjamin Franklin saying that "it is better a hundred guilty persons should escape than one innocent person should suffer" (Bigelow 1904). These statements have become a common convention, implying that people follow the norm not to apply sanctions if the information is very noisy.

The reason for abstaining from sanctioning in a noisy information system derives from a common understanding: a sanction regime may cause two distinct types of errors, "Type I" errors, which is the case when innocent defendants are found guilty, and "Type II" errors, when guilty defendants escape punishment (see, e.g., Ehrlich 1982; Polinsky & Shavell 1989; Miceli 1991). According to the common understanding, the social damage of a regime that minimizes "Type I" errors at the cost of "Type II" errors is substantially lower than the damage of a regime that allows for "Type I" errors in order to avoid "Type II" errors. Although various authors (recently, Blume 2008; Feess & Wohlschlegel 2009) have emphasized the costs of "Type I" errors, there is currently little evidence about the size of these costs.[2] The purpose of this article is to explore (1) whether people in fact exhibit a decline in their willingness to execute punishment when they are uncertain about others' behavior, and (2) whether social welfare in a sanctioning regime (with "Type I" and "Type II" errors) is in fact superior to a regime where no sanction possibilities are available, that is, where "Type I" errors cannot occur but "Type II" errors are frequent (without sanctions all defectors remain unpunished).

Experimental social dilemma games are an ideal workhorse for this class of questions, as people interact in groups in such a way that cooperation is optimal from a welfare

---

[1]Compare, e.g., Png (1986), Rubinfeld and Sappington (1987), Andreoni (1991), Volokh (1997), Rachlinski and Jourden (2003), Eisenberg et al. (2005), Lando (2006) Polinsky and Shavell (2007), and Eisenberg and Hans (forthcoming).

[2]Some economic scholars have explored the optimal degree of noisiness for sanctioning relevant information, given that one can quantify the social costs of "Type I" and "Type II" errors (see, e.g., Kaplow & Shavell 1994; Polinsky & Shavell 2000, 2007; Lando 2009; Rizzolli & Saraceno 2009).

perspective, but defection is rational for each individual. Among other mechanisms, like social norms and habits, sanctions serve as an important mechanism in maintaining social cooperation, as they offer participants the ability to enforce pro-social behavior by sanctioning defection. Previous research has drawn special attention to decentralized sanctioning, where subjects can spend money to distribute points that reduce other group members' incomes. Fehr and Gächter (2000, 2002) show that subjects mainly use the punishment option to discipline free-riders, allowing the group to attain high contributions and therefore producing efficient outcomes.[3] Results suggest that decentralized sanctions are a robust mechanism for stabilizing cooperation in anonymous groups. Yet, this result is obtained in a system with completely accurate information about the cooperation rates of all group members. To the best of our knowledge, we are the first to vary systematically the accuracy of contribution signals in a public goods game with a sanctioning mechanism.[4] The crucial question is how people behave if they receive noisy information about others' behavior. One could claim that, following the prevalent idea of a common rationale to avoid "Type I" errors, subjects abstain from applying sanctions. As a consequence, the group's total welfare declines if there is noisy information: defectors will not be disciplined by sanctions, meaning that cooperation cannot be improved.

In fact, our results are worse. We find that a large degree of noise does not discourage punishment. Our evidence even suggests that the more noise we introduce, the more punishment subjects apply. Particularly, the introduction of noise has two opposing effects: (1) noise increases the frequency of punishment acts, but (2) it decreases the intensity of punishment for a specific punishment act. Subjects do not differ systematically with respect to their response to received punishment in the noisy information regime and in the accurate information regime. As a consequence, under noise, defectors are mildly sanctioned and adjust their behavior only slightly (although severe punishment would lead to stronger corrections), and many cooperators are punished. The overall welfare assessment of punishment under noisy information is devastating. Compared to the punishment regime under accurate information, the introduction of a minor degree of noise already decreases welfare substantially; welfare is similar to that obtained in a game without a punishment mechanism. Introducing a major degree of noise yields welfare below that obtained in the game without punishment. This result is remarkable, as people could simply choose not to make use of punishment.

Our results have important policy implications. Even though there may be circumstances in reality with perfect information, in the vast majority of legal cases information is unavailable or prohibitively costly to obtain. Cases brought to court typically involve

---

[3]Earlier studies on sanctions in social dilemma games are Yamagishi (1986) and Ostrom et al. (1992). Herrmann et al. (2008) show, however, that the positive effect of the punishment option on contributions is not ubiquitous. They report data from a cross-cultural experiment showing that the effectiveness of the punishment option depends on cultural factors. Other studies providing a more ambiguous picture of punishment behavior are Miller (1999), Nikiforakis (2008), and Abbink and Herrmann (2009).

[4]Loosely related is a study by Fatas et al. (2010), who study the effect of a central sanctioning mechanism that punishes arbitrary subjects dependent on a group's joint contributions. Levati et al. (2009) study the effect of uncertainty regarding the marginal benefits of the public good.

uncertainty with respect to crucial facts. Not surprisingly, rules of legal procedure are inherently based on the incidence of error (e.g., Shavell 2004:451; Wistrich et al. 2005) and the law in general respects the fact that some information is private (e.g., Baird et al. 2003:79). Our findings suggest that regulators may be well advised to restrict sanctions under uncertainty, for example, through high standards of proof, even if that implies not offering any sanctioning mechanism in a noisy environment.

There are many kinds of sanctions, including reports of criminal behavior to the police, economic sanctions under international law, private lawsuits, and so forth, where our findings play an important role. One field that reflects our results nicely is international law. It deals with social dilemma settings (pollution, use of natural resources, nuclear activities, etc.), where information is typically difficult or impossible to obtain due to the nature of the issues and because of state sovereignty (Shaw 2008). Sanctions are often decentralized and costly to both parties (Guzman 2008). The fact that international law restricts the application and the magnitude of such sanctions and the fact that many treaties include information obligations follow the rationale of our findings. Even though some features have historically evolved for different reasons (Brownlie 2008; Shaw 2008), our results suggest that they may be beneficial from a welfare perspective. For similar reasons, states often prefer soft law, where potential sanctions are lower, over hard law, when they enter into an agreement (Guzman & Meyer 2009). Our research also provides evidence for the intuition that informal, that is, decentralized, sanctions work best if the parties can observe each others' actions (Posner 2007) and the intuition that uncertainty is a major barrier to inducing cooperation through sanctions in international law (cf. Abbott & Snidal 2004:62; Hirsch 2004:178).

The problem of erroneous sanctions is also important in criminal law. Unlike in our experiment, sanctions are usually carried out by a centralized institution. Still, judges and juries are part of the community and, as such, operate in a public goods environment. Here, convictions require proof beyond reasonable doubt for every fact necessary in constituting a crime (U.S. Supreme Court 1970). Considering the enormous social costs of "Type I" errors, and, on the other hand, the people's willingness to impose sanctions even under high uncertainty, rules of evidence that require substantial information are socially optimal.

Along a similar line of arguments, the costs of sanctions on cooperative behavior have also been studied in tax law. Here, the basis for a tax morale is a psychological contract (different from a legal contract to which the parties agreed ex ante), suggesting that disrespectful treatment by tax authorities erodes an intrinsic motivation to pay the dues. From this point of view, it is wise to give taxpayers the "benefit of doubt" (Feld & Frey 2002, 2003).

All these applications share an interesting feature: decisions on sanctions have to be taken under incomplete information, and higher uncertainty is supposed to discourage the use or reduce the intensity of the sanction. In this article, we analyze whether we can replicate this feature in a world with different, experimentally controlled degrees of uncertainty. We proceed as follows. We begin by describing our experimental design, which exposes subjects to an environment where they can cooperate in providing a public good. Treatment conditions vary as to whether subjects are able to impose costly sanctions on others and as to the accuracy of information with respect to others' cooperation rates. We

then discuss some expectations regarding subjects' behavior vis-à-vis different degrees of information accuracy. Finally, we present our experimental results and conclude with a discussion of policy implications.

II. Design

Our experimental tool is the standard voluntary contribution mechanism (VCM) with and without decentralized punishment. This design has been widely tested (for an overview, see Zelmer 2003) and it allows investigating cooperation and punishment behavior and comparing the efficiency of different punishment institutions. It is not tailored to a specific application in criminal or international law, but it is a framework that incorporates important features of many legally relevant situations.

We analyze behavior in a standard repeated VCM game with four players and 10 periods. The group composition remains constant over the 10 periods (partner design). At the beginning of each period, each player receives an endowment of 20 ECU (experimental currency units). Players simultaneously choose how many ECU to contribute to the public good, g_i, with $g_i \in \{0, 1, 2, \ldots, 20\}$. Each ECU contributed to the public good yields a benefit of 0.4 ECU (the marginal per-capita return) to *every* player in the group.

After the contributions are made, each player receives a signal s_j $(j \neq i)$ about the contributions of each other player in the group, such that:

$$s_j = \begin{cases} g_j \text{ with probability } \lambda, \\ \tilde{g}_j \text{ with probability } (1-\lambda), \end{cases} \tag{1}$$

where $\tilde{g}_j$ is an independent random draw from $\{0, 1, 2, \ldots, 20\} \setminus \{g_j\}$, all numbers with equal probability. Thus, for the contribution signals of the other three players, there is an independent random draw for each player, determining whether he or she receives the accurate signal, and if not, another independent draw that determines a random number to display. Each number (except g_i) is equally likely to appear if the signal does not correspond to the true contribution. For example, suppose λ equals 0.5 and Player 1 contributes 10 ECU. There is a probability of 50 percent that Player 2 sees the signal "10 ECU" for Player 1's contribution, while with a probability of 50 percent Player 2 sees a randomly picked number between 0 and 20, except 10 (e.g., "3 ECU"). The same is true for Player 3 and Player 4 regarding Player 1's contribution, such that Players 2, 3, and 4 may see the same or different true or wrong signals.

The labels "Player 1," "Player 2," and so forth are randomly assigned anew to players at the beginning of each period, making the identification of other players across periods impossible.

In the game without the punishment option, players are then informed about their earnings in the period and proceed to the next period. In the game with the punishment option, players enter a second stage. Here, they can punish the three other players. For this purpose, each player receives an extra endowment of 10 ECU in the second stage of every

period that cannot be contributed to the public good.[5] Each punishment point assigned to another player leads to a deduction of three ECUs from the punished player's account, but also reduces the punisher's income by one ECU. In sum, each player can spend up to 10 ECUs on (total) punishment in each period. ECUs not spent on punishment are credited to the particular player's account. Denoting punishment points that player i assigns to player
j as p_i^j, it follows $\sum_j p_i^j \leq 10$. Player i's payoff in a given period is:

$$\pi_i = 20 - g_i + 0.4\sum_j g_j + \left(10 - \sum_{j \neq i} p_i^j\right) - 3\sum_{j \neq i} p_j^i. \quad (2)$$

After each period, players learn their own payoff and the points they have received (however, they receive no detailed information on who distributed points). Players then proceed to the next period; payoffs accrue over 10 periods. All parameters, the signal technology, and payoff functions are common knowledge.

We apply four treatment conditions, three with a sanctioning mechanism and one without sanctioning.

- In the P/1 treatment, subjects receive accurate signals about the other group members' contributions ($\lambda = 1$) and may use the sanctioning mechanism.
- In the P/.9 treatment, the signal is accurate in 90 percent of the cases. In 10 percent of the cases, the signal does not correspond to the contribution of the other group member ($\lambda = 0.9$). After receiving information on the contributions, subjects may use the sanctioning mechanism.
- In the P/.5 treatment, the signal is accurate in only 50 percent of the cases ($\lambda = 0.5$). After receiving information on the contributions, subjects may use the sanctioning mechanism.
- In the N/.5 treatment, the signal is accurate in only 50 percent of the cases ($\lambda = 0.5$), and there is no sanctioning mechanism available.

For comparison, we use data from an experiment by Herrmann et al. (2008), who conducted a VCM game with identical parameters, in the same lab, and with the same subject pool, without punishment and accurate signals (denoted as N/1).

We ran a total of eight sessions with 48 groups (192 subjects), providing us with 12 independent observations per treatment condition. Each subject participated in only one treatment condition; none of the subjects had previously participated in a public goods experiment. The experiments were conducted at the laboratory for economic experiments (EconLab) at the University of Bonn in January to March 2009 with mostly undergraduate

---

[5]For a clean comparison between games with and without punishment option, we also pay the extra endowment of 10 ECU in the public goods game without punishment. By introducing an extra endowment for punishment, we depart slightly from the design found in the literature. We do so in order to avoid the following problem: for inaccurate signals there is uncertainty about the earnings from the public good at the punishment stage. The degree of this uncertainty depends on our treatment variable (signal quality). Separating money for punishment expenditures from the earnings of the public goods game is necessary to avoid undesired effects of our treatment variation on punishment behavior.

students from various fields.[6] Once all subjects were seated, the written instructions were handed to them before the experimenter read them out aloud.[7] Subjects were given the opportunity to ask questions (in private). Before the experiment started, subjects had to answer a set of control questions.[8] A session lasted about 60 minutes. Payoffs were converted at an exchange rate of 1 Euro per 40 ECUs. Subjects earned on average 13.67 Euro[9] (standard deviation 1.30 Euro), including a show-up fee of 2.50 Euro.

III. Expected Behavior

Assuming common knowledge of rationality and selfish preferences, the unique subgame perfect Nash equilibrium is that no player contributes to the public good. The reason is simple: each ECU contributed yields 0.4 ECU but costs 1 ECU. If the game is played for a finite number of periods, the rationale remains unchanged; reasons like reputation building at the beginning of the sequence do not matter from a theoretical point of view. Since no player contributes in the last period (in this period, reputation building is irrelevant), it is also rational not to contribute in the second to last period, and so on until the very first period. Irrespective of the other group members' decisions, not contributing maximizes payoffs. However, the group earns *four* times 0.4 for each ECU contributed. Therefore, the social welfare of the group (defined as the sum of payoffs of all group members) is maximized if all players fully contribute. Hence, the VCM game provides us with an individual measure for cooperation and allows us to investigate the efficiency of group outcomes.

What changes are there if we introduce sanctions? Given that punishment is costly for the punisher, norm enforcement by punishment itself is a public good: the entire group participates in the benefits stemming from the players who punish, while the punisher bears the costs alone. This design reflects the fact that enforcement of many legal rights is time consuming and not profitable from a purely monetary point of view. Hence, under standard assumptions, no player will exert punishment and contributions will be the same as in the game without sanctions.

There is ample experimental evidence that theoretical predictions under standard assumptions are a poor description of actual behavior. Subjects contribute to the public good, subjects make use of the punishment option, defectors receive punishment points, and cooperators sometimes receive punishment, too. Contributions decline over time if sanctioning is not available, but remain stable or increase if sanctioning is possible (see,

---

[6]Four percent of participants were nonstudents, 52 percent of participants were females, and age ranged between 16 and 47 (median 22). The experiment was computerized and programmed in zTree (Fischbacher 2007); we used ORSEE (Greiner 2004) for recruiting.

[7]Instructions were adapted from Hermann et al. (2008); a translated English version is available from the authors upon request.

[8]Control questions are available from the authors upon request.

[9]13.67 Euro corresponds to US$ 18.70 in February 2010.

e.g., Herrmann et al. 2008). All existing results, however, are established only under accurate information about other subjects' contributions.

How should inaccurate information about other players' contributions influence behavior? The most direct effect is certainly the effect on punishment. In treatments with punishment, noise makes separating defectors from cooperative subjects more difficult. Previous evidence strongly suggests that the targeted subjects' contribution is the major determinant of punishment. Most of the studies find that punishment is predominantly directed toward defectors (Fehr & Gächter 2000, 2002).[10] If punishment is used to enforce cooperation norms, subjects should become more reluctant to use the punishment option if there is the danger of erroneous punishment, that is, "Type I" errors. We thus hypothesize that the use of punishment decreases with decreasing information accuracy, that is, most punishment should be observed in P/1. We expect less punishment in P/.9, and the least amount in P/.5, where the signal is largely uninformative.

How do subjects respond to punishment? Although punishment is predominantly directed to defectors, some punishment is also directed toward cooperators, so-called antisocial punishment. Typically, antisocial punishment leads to a substantial decline in contributions from the targeted subject in subsequent periods. The crucial question is whether noise increases this damage due to an increased number of "Type I" errors.[11]

One can claim that punishment loses its legitimacy, meaning that subjects who receive punishment points in these cases cease to respond, as punishment is noisy. The victims of antisocial punishment may thus take this into account, suggesting that the social damage due to "Type I" errors is less severe than under accurate information. However, there might be more instances of erroneous punishment and, therefore, of antisocial punishment under noise. It is a priori unclear whether overall social damage due to "Type I" errors increases due to decreasing information accuracy.

Punishment is not the only channel, however, through which noise about the contribution information might affect cooperation. It is unclear whether less accurate information makes subjects more optimistic or more pessimistic about other subjects' contributions. Previous evidence suggests that a large fraction of the subjects can be characterized as conditionally cooperative (Fischbacher et al. 2001), that is, subjects who contribute only if they expect others to do so as well. Subjects in repeated public goods settings can use others' behavior in the previous period to form beliefs about their contributions in the current period. The fact that other subjects use the information about a subject's contributions introduces an incentive to signal cooperative behavior: subjects who expect others to be conditionally cooperative have a strategic incentive for choosing high contributions early in the game to induce other subjects to contribute. Noise may weaken

---

[10]Recent studies (e.g., Cinyabuguma et al. 2006) investigate punishment directed toward cooperators. Gächter et al. (2005) and Herrmann et al. (2008) show that the degree of such "antisocial punishment" is decisive with respect to the efficiency of the punishment mechanism in establishing cooperative results. The latter study also reports data from Bonn, where our experiments took place. In this subject pool, antisocial punishment is of little importance.

[11]Of course, given accurate information, antisocial punishment is not an error from the individual perspective. Reasons like spite or simply the joy of destruction may motivate this type of punishment. However, from the perspective of the entire group or of a social planner, this is erroneous punishment.

the strategic incentive for initial contributions, as the high cooperation signal is distorted. To identify effects of signal distortion on contributions, we compare the treatments without punishment opportunities (i.e., N/1 against N/.5).

Finally, a comparison of the P/.5 and the N/.5 treatment will allow us to test for the welfare implications of sanctioning under noisy information at its extreme. In other words, we analyze whether the social damage of a regime that minimizes "Type I" errors at the cost of "Type II" errors leads to superior social welfare than a regime that allows for "Type I" errors in order to avoid "Type II" errors. More precisely, a comparison of the two treatments answers the question: Does the regime N/.5, which rules out "Type I" errors by construction, but allows for "Type II" errors (due to the absence of any sanctioning mechanism), lead to more efficient outcomes than the regime P/.5, which is potentially subject to both types of errors? The answer to this question is a priori unclear; our experimental analysis will examine this question more closely.

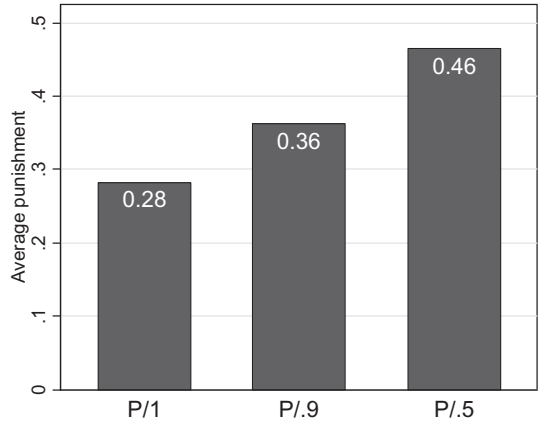
IV. Results

In the first subsection, we investigate how the use of the punishment option depends on the accuracy of the signal. Then we analyze our subjects' responses to received punishment and see how average contributions react to our treatment variables. Finally, we analyze overall welfare in the different treatment conditions.

A. Punishment Despite Reasonable Doubt

According to our hypotheses, there should be less punishment in treatments with higher degrees of noise. Figure 1 shows the average number of punishment points in the three

Figure 1: Average punishment points assigned over all periods by treatment condition.



Note: Average punishment is defined as the average number of punishment points in the particular treatment condition of the game.

treatments with punishment and differing signal accuracy. The result is surprising. Contrary to our hypothesis, higher noise leads to more punishment. On average, subjects distribute 0.46 punishment points per occasion[12] in P/.5, 0.36 points in P/.9, and 0.28 in P/1.[13] A conservative test based on 12 independent group averages in each treatment shows that the difference between P/1 and P/.5 is significant at $p = 0.043$ (Wilcoxon rank-sum test, two-sided). The comparisons between P/.9 and the other two treatments are not significant.

Punishment expenditures thus increase with noise. A closer look at the punishment decisions reveals that the introduction of noise has two opposing effects: (1) noise increases the frequency of punishment acts, but (2) it decreases the intensity of punishment for a specific punishment act. As shown in Figure 1, the former effect is clearly stronger, producing an overall increase of punishment with increasing noise. With respect to (1), we find that in 12.8 percent of all occasions subjects punish in P/1, while in 17.5 percent in P/.9, and in 25.2 percent in P/.5. The difference between P/1 and P/.5 is significant at $p = 0.024$ (Wilcoxon rank-sum test, two-sided), while the comparisons between P/.9 and the other two treatments are not significant. With respect to (2), we find that subjects—given that they punish—distribute on average 2.20 points in P/1 per occasion, while 2.07 points in P/.9, and 1.84 points in P/.5. The difference between P/1 and P/.5 is weakly significant ($p = 0.053$).

In a next step, we apply regression analyses in order to determine in more detail the determinants of the punishment decision. Table 1 shows Tobit estimates with the punishment decision as dependent variable. We report robust standard errors in parentheses, clustered on the group level. In Model 1, we use dummies for the two treatments P/.9 and P/.5 and the variable PERIOD to identify time effects. The dummy for P/.5 is significantly positive, indicating that punishment was stronger in the case where the signal was very unreliable. In Model 2, we include the difference between the signal and the punisher's contribution, $s_j - g_i$, denoted as SIGNAL-CONTRIBUTION. The literature identified the difference between a punisher's contribution and that of the punished subject as an important determinant of punishment behavior (see, e.g., Herrmann et al. 2008). We allow for different slopes for the cases where the signal is higher and lower than the punisher's contribution by introducing the variable SIGNAL-CONTRIBUTION IF > 0 for positive deviations: the variable equals SIGNAL-CONTRIBUTION if $s_j > g_i$, and is zero otherwise. The

---

[12]An occasion is a bilateral relation between two subjects in a period. Thus, a subject makes three punishment decisions in every period.

[13]Recall that we depart slightly from the standard design of the public goods game with punishment by introducing the extra endowment of 10 ECU and restricting total expenditures for punishment to 10 ECU. To test whether this design change affects punishment behavior, we compare data from our P/1 treatment to data by Herrmann et al. (2008), who conducted experiments in the standard design in the same subject pool. In their data, subjects distribute on average 0.38 punishment points per occasion, while we observe 0.28 points in our P/1 treatment. This difference is insignificant ($p = 0.714$, Wilcoxon rank-sum test, two-sided). If we pool their data with ours and run regressions as reported in Table 1, we are not able to find significant differences, either on the level or on the effect of the observed contribution differences. We conclude that the extra endowment does not systematically influence punishment behavior.

Table 1: Tobit Estimates for the Punishment Decision

	Dependent Variable: Punishment Points		
	Model 1	*Model 2*	*Model 3*
P/.9	0.628	–0.053	0.063
	(0.703)	(0.578)	(0.626)
P/.5	1.388**	0.249	1.952***
	(0.624)	(0.584)	(0.656)
Period	–0.124***	–0.091***	–0.078***
	(0.037)	(0.033)	(0.030)
SIGNAL-CONTRIBUTION		–0.304***	–0.474***
		(0.035)	(0.048)
SIGNAL-CONTRIBUTION IF > 0		0.388***	0.652***
		(0.067)	(0.071)
SIGNAL-CONTRIBUTION X P/.9			0.133**
			(0.059)
SIGNAL-CONTRIBUTION IF > 0 X P/.9			–0.118
			(0.098)
SIGNAL-CONTRIBUTION X P/.5			0.319***
			(0.065)
SIGNAL-CONTRIBUTION IF > 0 X P/.5			–0.586***
			(0.092)
Constant	–3.086***	–3.340***	–3.829***
	(0.551)	(0.469)	(0.559)
Standard error of estimate	3.422	2.907	2.804
F	4.920***	20.161***	29.390***
Pseudo R^2	0.012	0.091	0.110
N	4,320	4,320	4,320

NOTE: Tobit estimates for the punishment decision. Dependent variable is the number of punishment points. The dummy variables P/.9 and P/.5 indicate data that come from the corresponding treatment conditions. PERIOD identifies the period at which punishment occurred, SIGNAL-CONTRIBUTION the difference between the signal and the punisher's contribution. The variable SIGNAL-CONTRIBUTION IF > 0 equals SIGNAL-CONTRIBUTION if the signal is larger than the punisher's contribution and zero otherwise. Interaction effects between variables are indicated by "x." Robust standard errors, clustered on the group level, in parentheses. Summary statistics: standard error of the estimate; the *F* test statistic; pseudo R^2 reports the goodness of fit for the models; *N* denotes the number of observations. ***: significant at $p < 0.01$, **: at $p < 0.05$; *: at $p < 0.1$.

treatment dummies are insignificant once we control for the deviations in the contributions. The deviation variables have the expected signs: we observe a highly significant negative coefficient for negative deviations ($s_j < g_i$), that is, the lower the signal, the higher the punishment. We find evidence for antisocial punishment for positive deviations, given a significantly positive coefficient.[14]

However, the dependence of the punishment decision on the signal-contribution difference is likely to depend on the noise of the signal. In Model 3 of Table 1, we allow for differences in the reaction to the signal between the three treatments. We introduce interaction terms for the treatment dummies and the two measures for deviation. In case of

[14]The effect represents the sum of both estimated coefficients (that is, $-0.304 + 0.388 = 0.084$), which, jointly tested, is significant ($p = 0.042$).

negative deviations ($s_j < g_i$), both treatments with noisy signals have significantly less steep slopes, indicating that punishment is less strongly connected to the deviation between the signal and the own contribution. In other words, if the signal indicates that a subject is likely to be a free-rider, punishment is weaker if signals are less accurate. In case of positive deviations, only the interaction term for the treatment with high noise is significant, which means that high positive deviations are punished less strongly in P/.5 compared to the treatment with the perfect signal. If we allow for different slopes, the treatment dummy for P/.5 becomes highly significant, suggesting that small deviations are punished much more strongly under a high degree of noise. As in Model 1, the highly significant coefficient of P/.5 indicates stronger punishment with higher degrees of noise.

To summarize our results about subjects' willingness to distribute points, contrary to our expectations, noise does not discourage punishment. Subjects do not seem to take "Type I" errors into account and in fact punish *despite reasonable doubt*! However, the connection between the signal and the punishment decision is substantially weakened. Our subjects thus do react to our treatment variation. However, noise does not discourage punishment, but simply leads to more unsystematic punishment.

B. The Reaction to Rightful and Wrongful Sanctions

Exploring the response to received punishment allows us to test whether punishment under noise stabilizes or erodes cooperative behavior. Experimental evidence suggests that under perfect information, punishment maintains or even enhances cooperation due to the fact that free-riders increase their contribution when being punished. The reaction to received punishment is much less clear under noise because a subject never knows whether the punishment was deliberate or due to a false signal. To analyze the effect of received punishment, we run ordinary least squared (OLS) regressions for the difference in a subject's contributions between two consecutive periods as the dependent variable. In particular, the dependent variable is the difference between the CONTRIBUTION in $t+1$ and the CONTRIBUTION in t ($g_i^{t+1} - g_i^t$). Thus, a negative difference indicates a decrease in contributions, while a positive difference indicates an increase in contributions.

As before, we use the dummy variables P/.9 and P/.5 to identify the effect of noisy information. Furthermore, let us define the variable RECEIVED PUNISHMENT as the number of points subject i receives in period t. This variable measures the reaction to received punishment points. Interaction terms with P/.9 and P/.5 identify differences across treatment conditions. Finally, we introduce two control variables to disentangle the effect of punishment from other variables that influence contribution decisions. First, we include the variable PERIOD to identify time effects. Second, in order to control for peer effects, we measure the average of the contributions by all other subjects in t by the variable CONTRIBUTION OTHERS. Thus the variable indicates positive effects of observing other subjects contributing to the public good on own contribution decisions.

The reaction to received punishment is likely to differ between subjects with a high contribution and those with a low contribution. We therefore estimate two regression models, one for contribution decisions where the subject contributed less than the average in period t; and one for contribution decisions where the subject contributed the average

Table 2: Estimates for the Response to Received Punishment

	Dependent Variable: Contribution (t + 1) − Contribution (t)	
Contributed in t	*Less than Average*	*More or Equal to Average*
P/.9	0.374	–0.501*
	(1.006)	(0.277)
P/.5	0.118	–1.426**
	(0.995)	(0.558)
RECEIVED PUNISHMENT	0.724***	–0.465***
	(0.172)	(0.142)
RECEIVED PUNISHMENT X P/.9	–0.068	–0.017
	(0.276)	(0.316)
RECEIVED PUNISHMENT X P/.5	–0.156	0.286
	(0.329)	(0.291)
PERIOD	–0.227**	–0.232***
	(0.092)	(0.050)
CONTRIBUTION OTHERS	–0.099	0.084*
	(0.066)	(0.050)
Constant	3.250**	–0.596
	(1.477)	(0.961)
$F(9,35)$	6.450***	10.713***
R^2	0.100	0.069
N	462	834

NOTE: OLS regression for the change in contribution from t to $t+1$. The dependent variable is the difference between the contributions in period $t+1$ and t for a subject. The dummy variables P/.9 and P/.5 indicate data that come from the corresponding treatment conditions. PERIOD identifies period t, the variable RECEIVED PUNISHMENT is the sum of punishment points the particular subject received in period t. The variable CONTRIBUTION OTHERS equals the average contribution to the public good of all other subjects in the group in period t. Interaction effects between variables are indicated by "x." Robust standard errors, clustered on the group level, in parentheses. Summary statistics: the F test statistic; R^2 reports the goodness of fit for the models; N denotes the number of observations. ***: significant at $p < 0.01$, **: at $p < 0.05$; *: at $p < 0.1$.

or more than the average in period t. Thus, the separation into two models allows us to test whether the effect of punishment received as a free-rider differs from that received as a cooperator. The former we call pro-social punishment, the latter antisocial punishment.[15] The results of our estimations are summarized in Table 2. The first column reports the findings for received pro-social punishment, the second column for received antisocial punishment.

Our estimations show a number of interesting results: with respect to the response to punishment, we find a positive response to received pro-social punishment, whereas there is a negative response in terms of contributions to received antisocial punishment (as indicated by the significant positive and negative coefficients for RECEIVED PUNISHMENT). In contrast to punishment behavior, there seem to be no systematic differences in the

[15]Unlike before, the definition of pro- and antisocial punishment relies on the consequences, not on intentions. Judging whether punishment is intentionally antisocial or erroneous is difficult in the presence of noise because punished subjects receive no feedback on whether the punisher acted under accurate information. Therefore, a receiver's reference point for determining antisocial punishment (if there is one at all) is the average contribution.

response to punishment across treatment conditions. This result suggests that "Type I" errors of punishment cause substantial damages to social welfare, regardless of whether the information is perfect or imperfect.

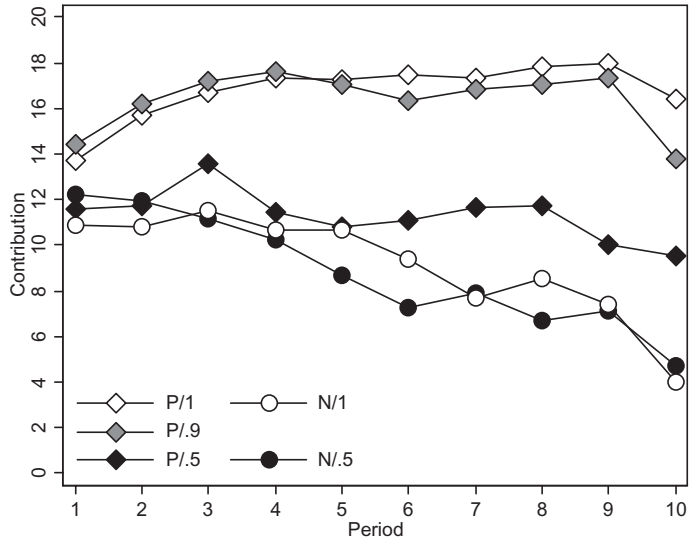
Overall, the significant negative coefficients for P/.9 and P/.5 show that contributions decline more strongly for those subjects who made average contributions or contributed more than the average if information is noisy. The variable CONTRIBUTION OTHERS remains insignificant for subjects who contributed less than the average in period *t*, while there is a weakly significant (positive) peer effect for subjects who contributed the average or more than the average. Finally, there is a significant negative time trend for the change in contribution, identified by the variable PERIOD.

To summarize our findings, we find the effect of "Type I" errors of punishment to be substantial and negative, and this effect does not seem to differ with respect to varying degrees of information accuracy.

C. Uncertainty and Contributions

Based on the results for punishment, we expect to find lower cooperation rates for increasing degrees of noise. Figure 2 shows the average contributions over the 10 periods in the five treatments. When subjects are perfectly informed about others' contributions, we confirm previous experimental studies and find significantly more cooperation if the punishment mechanism is available (16.8 in P/1 vs. 9.2 in N/1, $p < 0.001$, Wilcoxon rank-sum test, two-sided). This well-known result in the literature (Fehr & Gächter 2000,

Figure 2: Contributions over the *10* periods.



NOTE: Average contribution over the 10 periods and separated by treatment condition. Data of N/1 from Herrmann et al. (2008).

2002) does not hold, however, if we introduce noise. With a high degree of noise, punishment does not lead to significantly higher contributions when compared to the game without punishment but otherwise identical information conditions (11.3 in P/.5 vs. 8.8 in N/.5, $p = 0.133$). For mildly imperfect information, we find no significant difference in average contributions (16.8 in P/1 vs. 16.4 in P/.9, $p = 0.326$). A low degree of noise does not seem to affect cooperation. Punishment with accurate or nearly accurate information over contributions leads to very high levels of contributions (compare Figure 2).

On the other hand, we find a highly significant difference between the punishment treatments P/1 and P/.5 ($p < 0.001$). A high level of noise substantially harms the functioning of the punishment mechanism. However, punishment still appears to have some stabilizing force.[16]

In addition, the data allow us to investigate the effect of uncertainty on contributions in the absence of the punishment option. We find no significant differences in this case (9.2 in N/1 vs. 8.8 in N/.5, $p = 0.807$). As long as no punishment mechanism is available, the contribution rates are almost identical across treatments, suggesting that noise by itself has no effect. Thus, neither signal distortion nor its implication on the strategic incentives to contribute initially to the public good seems to influence cooperation rates essentially.

D. Welfare Consequences of Punishment Under Noisy Information

In the final step, we investigate how the availability of punishment under noise affects social welfare overall. In other words, we test whether introducing the risk of "Type I" and "Type II" errors for punishment harms efficiency. Efficiency is a linear function of contributions in the treatments without punishment, and the two treatments without punishment are therefore almost identical with regard to efficiency.[17] There are additional efficiency losses in treatments with punishment, due to received punishment and punishment expenditures.

P/1 is significantly more efficient than P/.5. Average earnings over the 10 periods are 36.7 in the former and 31.2 in the latter. This difference is highly significant ($p = 0.007$, Wilcoxon rank-sum test, two-sided). Thus, the reduction of information between the two treatments has real costs for the subjects. In relative terms, players in P/.5 realize only 10 percent of the maximum gains that could have been realized from the public good, while they realize 56 percent in P/1.[18] Comparing P/.5 and N/.5 shows that if noise is present, overall welfare is significantly lower when the sanctioning mechanism is available (31.2 vs.

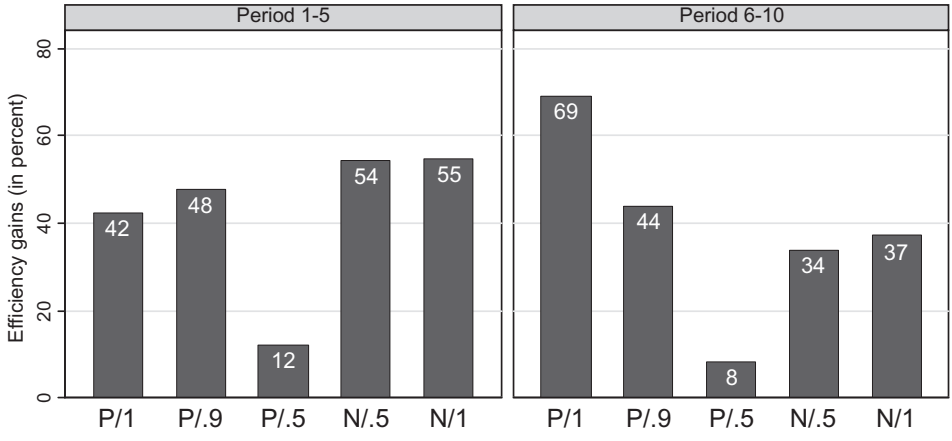
---

[16]Indeed, a simple OLS regression with the period as single independent variable results in no significant trend ($\beta = -0.24$ and $p = 0.178$) in P/.5. For two treatments without punishment, we observe a much stronger and highly significant negative trend ($p = 0.000$).

[17]Since participants received no extra endowment in N/1, we hypothetically add 10 ECU to their payoffs to compare results with other treatments.

[18]Maximal efficiency (100 percent) is obtained if all subjects earn 42 ECU (and, consequently, average earnings of 42 ECU), implying full contributions by all group members and no punishment at all. As minimal efficiency (0 percent), we define the outcome of the Nash equilibrium under standard assumptions, which is 30 ECU (in treatments with punishment, lower payoffs are possible, but did not occur in our experiment).

Figure 3: Realized efficiency gains in percent.



NOTE: Efficiency gains relative to a situation with zero contributions and no punishment (corresponding to 100 percent). Results are separated by treatment condition and by the first and second half of the experiment. Data of N/1 from Herrmann et al. (2008).

35.3, $p = 0.007$). In relative terms, there is an increase from 10 percent (P/.5) to 44 percent (N/.5). This result is striking, since participants could simply choose not to make use of the punishment mechanism.

Given that contributions decline in the treatments without punishment, we check whether this difference remains significant in later periods.[19] For this purpose, we compute relative efficiency gains for the first (Periods 1–5) and the second half (Periods 6–10) of the experiment separately in Figure 3. Comparing the perfect information settings over all 10 periods, welfare is higher when punishment is available, though insignificantly so ($p = 0.157$, Wilcoxon rank-sum test, two-sided). This insignificance is due to the fact that punishment does not yield immediate effects but needs some time to fully discipline noncooperators. Thus, a substantial amount of efficiency is lost in order to maintain and increase cooperation. We find significant differences in efficiency ($p = 0.008$) when punishment is available for the second half of the experiment, where the sanctioning mechanism enhances efficiency.

Finally, the efficiency obtained in the P/.9 treatment reveals an important result. Although information contains only a mild degree of noise, efficiency is hardly enhanced between the first and second halves of the experiment, in contrast to the P/1 treatment. As a consequence, efficiency in the second half is lower in P/.9 than in P/1 ($p = 0.072$) and does not differ significantly from that obtained in N/.5 and N/1 ($p = 0.162$, the two N treatments pooled). Even a very small amount of noise has profound implications for welfare consequences of punishment.

---

[19]Previous evidence shows that efficiency gains from punishment are often realized only in later periods of interactions (see, e.g., Gächter et al. 2008).

On the other hand, the efficiency gain in case of punishment with perfect information compared to the treatments without punishment is even stronger in the second half ($p = 0.002$, the two N treatments pooled). The efficiency loss under high degrees of noise relative to the treatments without punishment is persistent in the second half of the experiment ($p = 0.019$, the two N treatments pooled).

V. Discussion

Our experimental results show that noise crucially influences the effect of punishment on cooperation. The consequences are dramatic: with some nontrivial degree of noise, punishment cannot establish high cooperation levels; moreover, it decreases efficiency substantially! In this case, efficiency is even significantly lower than in a world without punishment mechanisms. This result is surprising, since people could simply choose not to make use of punishment. Even very little noise decreases efficiency in later periods when there exists a punishment option. Despite its negative implications, people are willing to punish and do so even under a high degree of noise. People spend a substantial amount on punishment, while cooperation is poorly maintained.

There are a number of examples from the law that correspond nicely to our findings, including international law (where states are increasingly treated as single players) (cf. Petersen 2009). When states enter into an agreement and decide on sanctions, they will balance the benefit of a formal treaty against its cost (Guzman 2008:141). Some of these costs arise because of wrongful sanctions. The issue is apparent in many fields of international law, including the use of nuclear weapons, the exploration of natural resources, pollution, and the like, where the parties typically receive only noisy information about the others' actions. As a result, international agreements often require the parties to provide information (Brown 2004) and typically include low sanctions or no sanctions at all (Guzman & Meyer 2009).

The fact that sanctions are restricted in international law (absent any agreement) accounts for both the scarcity of information available and for the decentralization of sanctions in a public goods environment. The cost of wrongful sanctions is one factor, among other considerations of self-restraints (Strüchler 2007; Gray 2008; Shaw 2008), that supports a prohibition on the use of force as it was pronounced in the Kellogg-Briand Pact of 1928 and amplified in the U.N. Charter after World War II. Acts of aggression mistakenly taken as a result of false information may erode cooperation. Due to the potentially severe consequences of sanctions under uncertainty, it is not surprising that governments have tried to increase the effectiveness of sanctions, including by "precise targeting" (Albright 1995). Past experience suggests that the use of force has not been very effective (Gray 2008:2).

This does not preclude a right to "self-defense" ("self-help"), but it emphasizes the importance of a high degree of accuracy. In the *Oil Platforms* case, the International Court of Justice made clear that the burden of proof is borne by the party that applies measures of self-defense (International Court of Justice 2003:189). Recent examples for sanctions under uncertainty include military actions after 9/11 (e.g., the war on Iraq was based on the

existence of weapons of mass destruction) (cf. Sandholz 2009) and the current conflict between North and South Korea (where a South Korean warship was hit by a torpedo, allegedly a North Korean one).

Under national legal systems, standards of proof serve as a similar limitation of errors. They range from a relatively high degree of information accuracy in criminal law to much lower requirements in civil procedures. Laws of civil procedure often require "clear and convincing evidence" or rely on a "preponderance of the evidence" standard (U.S. Supreme Court 1982:455 U.S. 745). Various quantifications of the standards of proof have been offered for both criminal offenses (Newman 1993; Volokh 1997; Tillers & Gottfried 2006) and civil procedure (Kaye 1982; Sanchirico 1997; Hay & Spier 1997). Rules of evidence that require substantial information may be socially optimal, given that people are willing to impose sanctions on others even under high degrees of noise. They prevent a dynamic that starts with an erroneous sanction due to false information, and leads to the erosion of pro-social behavior. To some extent, people intuitively try to account for this effect by imposing lower sanctions on others under noise if they decide to sanction. That is, even though people sanction more often under noise, they choose lower sanctions; vice versa, under certainty, people sanction others less often, but more strongly. Our results suggest that the intuition to condition the magnitude of a sanction on degree of noise poorly accounts for the effects of wrongful sanctions: either there is sufficient information, such that the optimal sanction should be imposed, or there is insufficient information, so that no sanction should be imposed. Typically, the law does not condition the magnitude of the sanction on the accuracy of the information, which is strongly supported by our results.[20] Whatever the optimal sanction and necessary information for a specific social dilemma is, intermediate sanctions are likely to be suboptimal. This is because sanctions, for a certain degree of noise, are simply inefficient for maintaining pro-social behavior. Standards of proof are a sensible tool for limiting the potentially devastating effects of wrongful sanctions.

Since cooperative behavior may be eroded, a system based on intrinsic motivations may be superior to a system with sanctions *in a noisy environment*. This is partly achieved through standards of proof (criminal law) and partly through limiting sanctions from the outset (international law). Policymakers have to take into account the quality of information on which punishment is based, as well as the corresponding welfare losses due to "Type I" errors. If sanctions are available, they need to be conditioned on substantial information accuracy, so that the social damage due to "Type I" errors is low.

References

Abbink, K., & B. Herrmann (2009) *Pointless Vendettas*. Available at <http://ssrn.com/abstract=1468452>.

Abbott, K., & D. Snidal (2004) "Pathways to International Cooperation," in E. Benvenisti & M. Hirsch, eds., *The Impact of International Law on International Cooperation*, pp. 50–84. Cambridge: Cambridge Univ. Press.

[20]Different standards of proof are used for criminal convictions and punitive damages (cf. Zipursky 2005; Lazer & Higgitt 2009; Markel 2009).

Albright, M. (1995) "Introduction: International Law Approaches the Twenty-First Century: A U.S. Perspective on Enforcement," 18 *Fordham International Law J.* 1595.

Andreoni, J. (1991) "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?," 22 *RANA J. of Economics* 385.

Baird, D., R. Gertner, & R. Picker (2003) *Game Theory and the Law*. Harvard: Harvard Univ. Press.

Bigelow, J., ed. (1904) *The Works of Benjamin Franklin*. New York.

Blackstone, W. (1765–1769) *Commentaries on the Law of England*. Available at <http://www.lonang.com/exlibris/blackstone>.

Blume, J. (2008) "The Dilemma of a Criminal Defendant with a Prior Record—Lessons from the Wrongfully Convicted," 5 *J. of Empirical Legal Studies* 477.

Brown, E. (2004) "Rethinking Compliance with International Law," in E. Benvenisti & M. Hirsch, eds., *The Impact of International Law on International Cooperation*, pp. 134–65. Cambridge: Cambridge Univ. Press.

Brownlie, I. (2008) *Principles of Public International Law*. Oxford: Oxford Univ. Press.

Cardozo, B. (1916) *People v. Galbo*, 218 N.Y. 283, 290, 112 N.E. 1041, 1044.

Cinyabuguma, M., T. Page, & L. Putterman (2006) "On Perverse and Second-Order Punishment in Public Goods Experiments with Decentralized Sanctioning," 9 *Experimental Economics* 265.

Ehrlich, I. (1982) "The Optimum Enforcement of Laws and the Concept of Justice: A Positive Analysis," 2 *International Rev. of Law & Economics* 3.

Eisenberg, T., P. Hannaford-Agor, V. Hans, N. Waters, G. T. Munsterman, S. Schwab, S., & W. Wells (2005) "Judge-Jury Agreement in Criminal Cases: A Partial Replication of Kalven & Zeisel's The American Jury," 2 *J. of Empirical Legal Studies* 171.

Eisenberg, T., & V. Hans (forthcoming) "Taking a Stand on Taking a Stand: The Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes," 94 *Cornell Law Rev.*

Fatas, E., A. J. Morales, & P. Ubeda (2010) "Blind Justice: An Experimental Analysis of Random Punishment in Team Production," 31 *J. of Economic Psychology* 358.

Feess E., & A. Wohlschlegel (2009) "Why Higher Punishment May Reduce Deterrence," 104 *Economic Letters* 69.

Fehr, E., & S. Gächter (2000) "Cooperation and Punishment in Public Goods Experiments," 90 *American Economic Rev.* 980.

—— (2002) "Altruistic Punishment in Humans," 415 *Nature* 137.

Feld, L., & B. Frey (2002) "Trust Breeds Trust: How Taxpayers are Treated," 3 *Economics of Governance* 87.

—— (2003) "Deterrence and Tax Morale: How Tax Administrations and Taxpayers Interact," 3/10 *OECD Papers* 1.

Fischbacher, U. (2007) "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," 10 *Experimental Economics* 171.

Fischbacher, U., S. Gächter, & E. Fehr (2001) "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," 71 *Economics Letters* 397.

Gächter, S., B. Herrmann, & C. Thöni (2005) "Cross-Cultural Differences in Norm Enforcement," 28 *Behavioral & Brain Sciences* 822.

Gächter, S., E. Renner, & M. Sefton (2008) "The Long-Run Benefits of Punishment," 322 *Science* 1510.

Gray, C. (2008) *International Law and the Use of Force*. Oxford: Oxford Univ. Press.

Greiner, B. (2004) "An Online Recruitment System for Economic Experiments," in K. Kremer & V. Macho, eds., *Forschung und wissenschaftliches Rechnen 2003, Bericht der Gesellschaft für wissenschaftlichen Dateverarbeitung Göttingen 63*, pp. 79–93.

Guzman, A. 2008 *How International Law Works*. Oxford: Oxford Univ. Press.

Guzman, A., & T. Meyer (2009) *Explaining Soft Law*. Available at <http://ssrn.com/abstract=1353444>.

Hay, B., & K. Spier (1997) "Burdens of Proof in Civil Litigation: An Economic Perspective," 26 *J. of Legal Studies* 413.

Herrmann, B., C. Thöni, & S. Gächter (2008) "Antisocial Punishment Across Societies," 319 *Science* 1362.

Hirsch, M. (2004) "Compliance with International Norms," in E. Benvenisti & M. Hirsch, eds., *The Impact of International Law on International Cooperation*, pp. 134–65. Cambridge: Cambridge Univ. Press.

International Court of Justice (2003) "Oil Platforms, Islamic Republic of Iran v. United States of America," 2003 *ICJ Report* 161.

Kaplow, L., & S. Shavell (1994) "Accuracy in the Determination of Liability," 37 *J. of Law & Economics* 1.

Kaye, D. (1982) "The Limits of the Preponderance of the Evidence Standard," 7 *American Bar Foundation Research J.* 487.

Lando, H. (2006) "Does Wrongful Conviction Lower Deterrence," 35 *J. of Legal Studies* 327.

—— (2009) "Prevention of Crime and the Optimal Standard of Proof in Criminal Law," 5 *Rev. of Law & Economics* 33.

Lazer, L., & J. Higgitt (2009) "Ascertaining the Burden of Proof for an Award for Punitive Damages in New York? Consult Your Local Appellate Division," 25 *Touro Law Rev.* 725.

Levati, M. V., A. Morone, & A. Fiore (2009) "Voluntary Contributions with Imperfect Information: An Experimental Study," 138 *Public Choice* 199.

Markel, D. (2009) "How Should Punitive Damages Work," 157 *Univ. of Pennsylvania Law Rev.* 1383.

Miceli, T. J. (1991) "Optimal Criminal Procedure: Fairness and Deterrence," 11 *International Rev. of Law & Economics* 3.

Miller, W. (1999) "In Defense of Revenge," in B. Hanawalt & D. Wallace, eds., *Medieval Crime and Social Control*, pp. 70–89. Minneapolis: Univ. of Minnesota Press.

Newman, J. (1993) "Beyond 'Reasonable Doubt'," 68 *New York Univ. Law Rev.* 979.

Nikiforakis, N. (2008) "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" 92 *J. of Public Economics* 91.

Ostrom, E., J. Walker, & R. Gardner (1992) "Covenants With and Without a Sword: Self-Governance is Possible," 86 *American Political Science Rev.* 404.

Petersen, N. (2009) "Rational Choice or Deliberation? Customary International Law Between Coordination and Constitutionalization," 165 *J. of Institutional & Theoretical Economics* 71.

Png, I. P. L. (1986) "Optimal Subsidies and Damages in the Presence of Judicial Error," 6 *International Rev. of Law & Economics* 101.

Polinsky, A. M., & S. Shavell (1989) "Legal Errors, Litigation, and the Incentive to Obey the Law," 5 *J. of Law, Economics, & Organization* 99.

—— (2000) "The Economic Theory of Public Enforcement of Law," 38 *J. of Economic Literature* 45.

—— (2007) "The Theory of Public Enforcement of Law," in A. M. Polinsky & S. Shavell, eds., *Handbook of Law and Economics*. Amsterdam: Elsevier Science Publishing.

Posner, E. (2007) "Review of *The Limits of Leviathan: Contract Theory and the Enforcement of International Law*, by Robert E. Scott and Paul B. Stephan," 101 *American J. of International Law* 509.

Rachlinski, J., & F. Jourden (2003) "The Cognitive Components of Punishment," 88 *Cornell Law Rev.* 457.

Rizzolli, M., & M. Saraceno (2009) *Better X Guilty Persons Escape than that One Innocent Suffer*, Working Paper, University of Milan–Bicocca.

Rubinfeld, D., & D. Sappington (1987) "Efficient Awards and Standards of Proof in Judicial Proceedings," 18 *RAND J. of Economics* 308.

Sanchirico, C. W. (1997) "The Burden of Proof in Civil Litigation: A Simple Model of Mechanism Design," 17 *International Rev. of Law & Economics* 431.

Sandholz, W. (2009) "The Iraq War and International Law," in D. Armstrong, ed., *Routledge Handbook of International Law*. London: Routledge.

Shavell, S. (2004) *Foundations of Economic Analysis of Law*. Harvard: the United States of America.

Shaw, M. (2008) *International Law*. Cambridge: Cambridge Univ. Press.

Strüchler, N. (2007) *The Threat of Force in International Law*. Cambridge: Cambridge Univ. Press.

Tillers, P., & J. Gottfried (2006) "Case Comment—*United States v. Copeland*, 396 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim that Proof Beyond a Reasonable Doubt is Unquantifiable?" 5 *Law, Probability & Risk* 135.

U.S. Supreme Court (1970) *In re Winship*, 397 U.S. 358.

—— (1982) *Santosky v. Kramer*, 455 U.S. 745.

Volokh, A. (1997) "N Guilty Men," 146 *Univ. of Pennsylvania Law Rev.* 173.

Wistrich, A., C. Guthrie, & J. Rachlinski (2005) "Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding," 153 *Univ. of Pennsylvania Law Rev.* 1251.

Yamagishi, T. (1986) "The Provision of a Sanctioning System as a Public Good," 51 *J. of Personality & Social Psychology* 110.

Zelmer, J. (2003) "Linear Public Goods: A Meta-Analysis," 6 *Experimental Economics* 299.

Zipursky, B. (2005) "A Theory of Punitive Damages," 84 *Texas Law Rev.* 105.