Punishment despite
Reasonable Doubt –
A Public Goods Experiment
with Uncertainty over
Contributions

Christoffel Grechenig
Andreas Nicklisch
Christian Thöni

MAX PLANCK SOCIETY

Punishment despite Reasonable Doubt
– A Public Goods Experiment with Uncertainty over Contributions

Christoffel Grechenig / Andreas Nicklisch / Christian Thöni

April 2010

Punishment Despite Reasonable Doubt

– A Public Goods Experiment with Uncertainty over Contributions

Kristoffel Grechenig

University of St. Gallen, Department of Law
Bodanstr. 4, 9000 St. gallen, Switzerland
+41 71 224 3006 ▪ Kristoffel.Grechenig@unisg.ch
&
Max Planck Institute for Research on Collective Goods
Kurt-Schumacher-Strasse 10, 53113 Bonn, Germany
+49 228 91416 51 ▪ Grechenig@coll.mpg.de

Andreas Nicklisch

Max Planck Institute for Research on Collective Goods
Kurt-Schumacher-Strasse 10, 53113 Bonn, Germany
+49 228 91416 79 ▪ Nicklisch@coll.mpg.de

Christian Thöni

University of St. Gallen, Department of Economics
Varnbüelstr. 14, 9000 St. gallen, Switzerland
+41 71 224 2548 ▪ Christian.Thoeni@unisg.ch

Abstract

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 good 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 others' behavior. We investigate punishment in a world with "reasonable doubt" about others' contributions. Interestingly, people reveal a high willingness to punish even if their information about cooperation rates is inaccurate, or noisy. If there is some non-trivial degree of noise, punishment (1) cannot maintain high contributions 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 can be optimal.

Introduction

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 to decide whether to impose sanctions would be aware of all relevant facts. Reality is much less perfect. Decision makers typically face imprecise, contradicting, 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, book 4, ch. 27) indicating that "the law holds, that it is better that ten guilty persons escape, than that one innocent suffer." In fact, this statement has become a common convention implying that people follow the norm not to apply sanctions if the information is very noisy. Sanctions include reports of criminal behavior to the police, economic sanctions under international law, private law suits, etc. The individual rationality 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. While various authors (recently Blume, 2008, Feess & Wohlschlegel, 2009) have emphasized

---

[1] Compare, e.g., Png (1986), Rubinfeld & Sappington (1987), Volokh (1997), Rachlinski & Foulden (2003), Eisenberg et al. (2005), Polinsky & Shavell (2007), Eisenberg & Hans (2009).

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 (a) whether people in fact exhibit a decline in their willingness to execute punishment given that they are uncertain about others' behavior, and (b) 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, i.e., where "type I" errors cannot occur but "type II" errors are frequent (without sanctions all defectors remain unpunished).

Experimental economics provides social dilemma games as ideal workhorses for this class of questions, as people interact in groups in such a way that cooperation is optimal from a welfare 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 distribute points which reduce both other group members' – and their own – income. Fehr & 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.

The result above, however, is obtained in a system with completely accurate information concerning the cooperation rates of all group members. To the best of our knowledge we are the first to systematically vary the accuracy of contribution signals in a public goods game with sanctioning mechanism.[4] The crucial question is how people behave if they are faced with 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.

---

[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).

[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 which shows the effectiveness of the punishment option depends on cultural factors.

[4] Loosely related is a study by Fatas et al. (2009), 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 about the marginal benefits of the public good.

In fact, our results are even 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. On the other hand, 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, defectors are mildly sanctioned and adjust their behavior only slightly (although severe punishment would lead to stronger corrections), whereas 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 equivalent to that obtained in a game without a punishment mechanism. Introducing a major degree of noise, finally, 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 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 not to offer any sanctioning mechanism at all if there is a substantial degree of noise. In other words, if installing a sanctioning system with or without punishment were an endogenous choice variable, the issue of information accuracy (or the costs for accurate information) would crucially affect the decision whether to opt for such a system or not.

One example which reflects our findings nicely is public 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). The fact that public international law is characterized by few (decentralized) sanctioning mechanisms and by many treaties with information obligations follows the rationale of our findings. Even though these features have historically evolved for different reasons (Brownlie, 2008, Shaw, 2008), our results suggest that they may be good from a welfare perspective.

Another example is criminal law where high standards of proof are required. Convictions under criminal law require proof beyond reasonable doubt for every fact necessary in constituting a crime (United States Supreme Court, 1970). Considering the enormous social costs of "type I" errors, and, on the other hand, people's willingness to impose sanctions even under high uncertainty, rules of evidence that require substantial information are socially optimal.

This article proceeds 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.

Design

Our experimental tool is the standard voluntary contribution mechanism (VCM) with and without decentralized punishment. This design has been widely tested (see Zelmer, 2003, for an overview) and it allows for both an efficiency analysis and for general applications to the law.

More precisely, we analyze behavior in a standard repeated VCM game with four players and ten periods. The group composition remains constant over the ten 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 $g_i \in \{0,1,2,...,20\}$. 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 $\{0,1,2,....,20\}\backslash\{g_j\}$, all numbers with equal probability. Thus each number (except g_i) is equally likely to appear if the signal does not

correspond to the actual contribution. The three signals are shown in a random sequence, making the identification of the other players 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 other players. For this purpose, each player receives an extra endowment of 10 ECU in every period.[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. Each player can spend up to 10 ECU on punishment. ECUs not spent on punishment are credited to the particular player's account. Denoting punishment points that player i assigns to player j as $p_i^j (j \neq i)$, it follows $\sum_j p_i^j \leq 10$. Player i's payoff in a given period, π_i, is

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

After each period, players learn their own payoff and the points they received (but get no detailed information on who distributed points). Players then proceed to the next period; payoffs accrue over ten 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 ten percent of the cases, the signal does not correspond to the contribution of the other group member ($\lambda=.9$). 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 ($\lambda=.5$). 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 ($\lambda=.5$), and there is no sanctioning mechanism available.

---

[5] By introducing an extra endowment for punishment, we depart slightly from the "standard setting" of this game. However, we avoid the following problem in doing so: The earnings from the public good (and, consequently, the average contribution of all players to the public good) are unknown at this stage of the game. The extra endowment prevents subjects from having to speculate about their earnings when choosing the amount of their income they want to spend for punishment points.

For comparison, we use data from an experiment by Hermann 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 8 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 good 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 students from various fields.[6] Once all subjects were seated, the written instructions were handed to them before the experimenter read them out loud.[7] Subjects were given the opportunity to ask any questions they might have privately. After questions had been answered individually, subjects had to solve a set of control questions.[8] A session lasted for about 60 minutes. Payoffs earned 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.

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.

---

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

[7] Instructions are adapted from Hermann et al. (2008); a translated English version is enclosed in Appendix A.

[8] Control questions are enclosed in Appendix B.

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

What changes 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 Hermann et al., 2008, for example). 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 towards 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, i.e., "type I" errors. We thus hypothesize that the use of punishment is decreasing for 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 towards 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 or not

---

[10] Recent studies (e.g., Cinyabuguma et al., 2006) investigate punishment directed towards 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.

noise increases this damage due to an increased number of "type I" errors.[11] One can claim that punishment loses its legitimacy, meaning that 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' previous period behavior 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 the strategic incentive for initial contributions, as the high cooperation signal is distorted. In order 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 following question: does the regime N/.5 that 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 that is potentially subject

---

[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 the one of a social planer, this is erroneous punishment.

to both types of errors? The answer to this question is a priori unclear; our experimental analysis will examine this question more closely.

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.

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 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. A conservative test based on twelve 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 (see below) reveals that the introduction of noise has two opposing effects: (i) noise increases the frequency of punishment acts, but (ii) 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.

---

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

Figure 1: *Average punishment points assigned over all periods by treatment condition.*
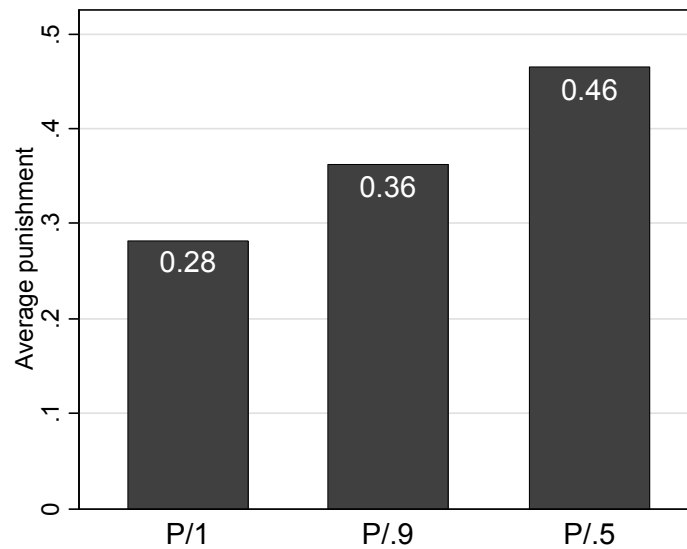
We apply regression analyses for a more detailed examination. Table 1 shows Tobit estimates with the punishment decision as dependent variable. We report robust standard errors in parentheses, clustered on the group level. In Model 1, we use dummies for the two treatments P/.9 and P/.5 and the variable *Period* to identify time effects. The dummy for the 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 *diff*. The literature identified the difference between a punisher's contribution and that of the punished subject as an important determinant of punishment behavior (see Herrmann et al., 2008, for example). We allow for different slopes for the cases where the signal is higher and lower than the punisher's contribution by introducing the variable *diff+* for positive deviations: the variable equals *diff* if $s_j > g_i$ but is zero otherwise. The 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.[13]

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, 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 negative deviations

---

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

($s_j < g_i$), 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.

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)
diff		-0.304***	-0.474***
		(0.035)	(0.048)
diff+		0.388***	0.652***
		(0.067)	(0.071)
diff x P/.9			0.133**
			(0.059)
diff+ x P/.9			-0.118
			(0.098)
diff x P/.5			0.319***
			(0.065)
diff+ 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 R²	0.012	0.091	0.110
N	4320	4320	4320

Note: Tobit estimates for the punishment decision. Robust standard errors, clustered on the group level in parentheses. Summary statistics: standard error of the estimate; the F-test statistic; Pseudo R² reports the goodness of fit for the models; N denotes the number of observations. ***: significant at $p < .01$, **: at $p < .05$; *: at $p < .1$.

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*! The connection between the signal and the punishment decision is substantially weakened, however. Our subjects thus do react to our treatment variation. However, noise does not discourage punishment, but simply leads to more unsystematic punishment.

The reaction to (mis)guided 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. In order to analyze the effect of received punishment, we run OLS regressions for the difference in a subject's contributions between two consecutive periods as the dependent variable. Particularly, let us define Δ_t as the contribution in period t less the contribution in period $t-1$, i.e., $\Delta_t = g_i^t - g_i^{t-1}$. Thus, a negative Δ_t indicates a decrease in contributions, while a positive Δ_t 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 *sum p rec* as the number of points subject i receives in period $t-1$. Hence, *sum p rec* measures the effect of punishment points received on contributions in the consecutive period. 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, in order to control for peer effects, we measure the sum of contributions by all other subjects in $t-1$ by the variable *sum c*. Thus the variable indicates positive effects of observing other subjects contributing to the public good on own contribution decisions. Second, we include the variable *Period* to identify time effects.

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 $t-1$; and one for contribution decisions where the subject contributed the average or more than the average in period $t-1$. 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.[14] 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.

Table 2: *Estimates for the response to received punishment*

	Dependent variable: Δ_t	
contributed in *t–1*:	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)
sum p rec	0.724***	−0.465***
	(0.172)	(0.142)
sum p rec x P/.9	−0.068	−0.017
	(0.276)	(0.316)
sum p rec x P/.5	−0.156	0.286
	(0.329)	(0.291)
sum c	−0.227**	−0.232***
	(0.092)	(0.050)
Period	−0.099	0.084*
	(0.066)	(0.050)
Constant	3.250**	−0.596
	(1.477)	(0.961)
$F_{(9,35)}$	6.450***	10.713***
R²	0.100	0.069
N	462	834

Note: OLS regression for the change in contribution from *t–1* to *t*. Robust standard errors, clustered on the group level in parentheses. Summary statistics: the F-test statistic; R² reports the goodness of fit for the models; N denotes the number of observations. ***: significant at $p < .01$, **: at $p < .05$; *: at $p < .1$.

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 *sum p rec*). In contrast to punishment behavior, there seems to be no systematic significant difference in this reaction across treatment conditions. This result suggests that "type I" errors of punishment cause

---

[14] 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.

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 contributed more or at the average if information is noisy. Interestingly, there is a significant negative peer effect (indicated by the coefficients of *sum c*), showing that subjects reduce their contribution the more others contributed to the public good in the previous period. Finally, there is only a weakly significant time trend for the change in contribution, again restricted to subjects who contributed more or at the average. Here, contributions decline towards the end of the experiment.

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 various degrees of information accuracy.

*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 ten 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 versus 9.2 in N/1, $p < 0.001$, Wilcoxon rank-sum test, two-sided). This well-known result in the literature (Fehr & Gächter, 2000, 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 versus 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 versus 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.[15]

In addition, the data allows 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 versus 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, signal distortion and its implication on the strategic incentives to contribute initially to the public good seem not to influence cooperation rates essentially.

Figure 2: *Contributions over the ten periods*
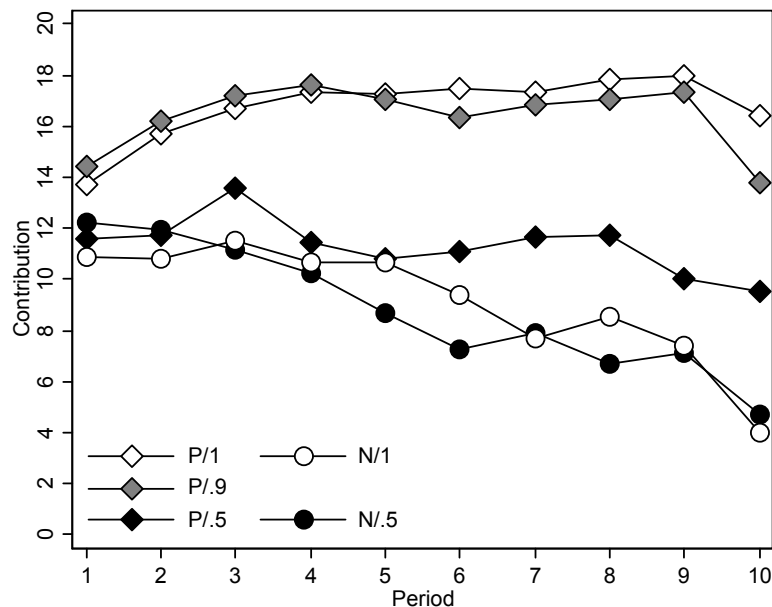
Note: Data of N/1 from Herrmann et al. (2008).

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 or not. Efficiency is a linear function of contributions in the treatments without punishment, and the two treatments without punishment are

---

[15] 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 ($p = 0.000$).

therefore almost identical with regard to efficiency.[16] 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 ten 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% of the maximum gains that could have been realized from the public good, while they realize 56% in P/1.[17] 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 versus 35.3, $p = 0.007$). In relative terms, there is an increase from 10% (P/.5) to 44% (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.[18] For this purpose, we compute relative efficiency gains for the first (period 1-5) and the second half (period 6-10) of the experiment separately in Figure 3. Comparing the perfect information settings over all ten 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 non-cooperators. 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 the second half 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

---

[16] Since participants received no extra endowment in N/1, we hypothetically add 10 ECU to their payoffs (as we actually did in the N/.5 treatment) to compare results with other treatments.

[17] Maximal efficiency (100%) 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%), 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).

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

pooled). Even a very small amount of noise has profound implications for welfare consequences of punishment.

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).

Figure 3: *Realized efficiency gains in percent*
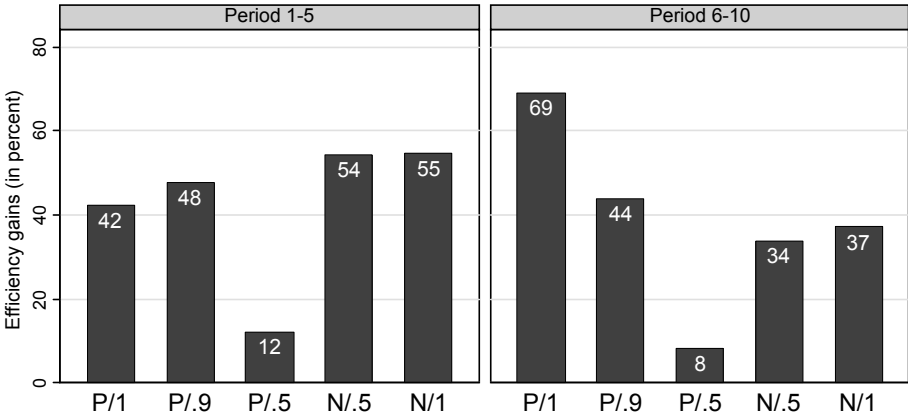
Note: Data of N/1 from Herrmann et al. (2008).

Conclusion

Our experimental results show that noise crucially influences the effect of punishment on cooperation. The consequences are dramatic: With some non-trivial degree of noise, punishment cannot maintain high cooperation levels; moreover, it decreases efficiency substantially! In this case, efficiency is even significantly lower than in a world without punishment mechanism. 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.

As we have already shown in the introduction, there is a number of examples from the law that correlate nicely with our findings. The fact that sanctions are restricted in public international law accounts for both the little information available to parties and the decentralization of sanctions in a public good environment. The underlying idea is similar to the prohibition of war pronounced in the Kellogg–Briand Pact of 1928 and amplified in the United Nations Charters after World War II. Acts of aggression mistakenly taken as a result of false information may erode cooperation. This development is almost unavoidable if the other parties' actions are not observable, unless there is a strict restriction of retributive behavior.

Under national legal systems, the standards of proof serve as a similar limitation of errors in sanctions. They range from a relatively high degree of information accuracy in criminal law to much lower requirements in civil procedures. Convictions under criminal law require proof beyond reasonable doubt of every fact necessary to constitute a crime (United States Supreme Court, 1970). Laws of civil procedure often require "clear and convincing evidence" or "preponderance of the evidence" (United States Supreme Court, 1982). Various quantifications of the standards of proof have been offered for both criminal offences (Newman, 1993, Tillers, 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. Standards of proof are a sensible tool for limiting potentially devastating effects of sanctions.

Overall, the legal implications of our findings are two-fold: if enforcement is based on noisy information, not offering sanctions might be better. Since cooperative behavior may be eroded, a system based on intrinsic motivations may be superior to a system with sanctions. In other words, in those instances where the initiation of a sanctioning system has to be chosen, decision makers have to take the quality of information on which punishment is based into account, as well as the corresponding welfare losses due to "type I" errors. If punishment is available, sanctions need to be conditioned on substantial information accuracy, so that the social damage due to "type I" errors is minimized.

the 4th Annual Conference on Empirical Legal Studies, the participants of the Ratio-Discussion-Group at the Max-Planck-Institute for Research on Collective Goods, and the participants of the Frey-Frey-Engel-Workshop.

References

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

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, Journal of Empirical Legal Studies 5, 477–505.

Brownlie, I. (2008), Principles of Public International Law, Oxford.

Cinyabuguma, M., Page, T., & Putterman, L. (2006), On perverse and second-order punishment in public goods experiments with decentralized sanctioning, Experimental Economics 9, 265–279.

Ehrlich, I. (1982), The optimum enforcement of laws and the concept of justice: A positive analysis, International Review of Law and Economics 2, 3–27.

Eisenberg, T., Hannaford-Agor, P., Hans, V., Waters, N, Munsterman, G.T., Schwab, S., & Wells, M. (2005), Judge-jury agreement in criminal cases: A partial replication of Kalven & Zeisel's The American jury, Journal of Empirical Legal Studies 2, 171–206.

Eisenberg, T., Hans, V. (2009), Taking a stand on taking a stand: The effect of a prior criminal record on the decision to testify and on trial outcomes, Cornell Law Review 94 (forthcoming).

Fatas, E., Morales, A.J., & Ubeda, P. (2009), Blind justice, Mimeo.

Feess E., & Wohlschlegel, A. (2009), Why higher punishment may reduce deterrence, Economic Letters 104, 69–71.

Fehr, E., & Gächter, S. (2000), Cooperation and punishment in public goods experiments, American Economic Review 90, 980–994.

Fehr, E., & Gächter, S. (2002), Altruistic punishment in humans, Nature 415, 137–140.

Fischbacher, U., (2007), z-Tree: Zurich toolbox for ready-made economic experiments, Experimental Economics 10, 171–178.

Fischbacher, U., Gächter, S., & Fehr, E. (2001), Are people conditionally cooperative? Evidence from a public goods experiment, Economics Letters 71, 397–404.

Gächter, S., Herrmann, B., & Thöni, C. (2005). Cross-cultural differences in norm enforcement. Behavioral and Brain Sciences 28, 822–823.

Gächter, S., Renner, E., & Sefton, M. (2008), The long-run benefits of punishment, Science 322, 1510.

Greiner, B. (2004), An online recruitment system for economic experiments, in: Kremer, K. & Macho, V. (eds.), Forschung und wissenschaftliches Rechnen 2003, Bericht der Gesellschaft für wissenschaftlichen Dateverarbeitung Göttingen 63, 79–93.

Hay, B., & Spier, K. (1997), Burdens of proof in civil litigation: An economic perspective, Journal of Legal Studies 26, 413–431.

Herrmann, B., Thöni, C., & Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.

Kaplow, L., & Shavell, S. (1994), Accuracy in the determination of liability, Journal of Law & Economics 37, 1-15.

Kaye, D. (1982), The limits of the preponderance of the evidence standard, American Bar Foundation Research Journal 7, 487–516.

Lando, H. (2009), Prevention of crime and the optimal standard of proof in criminal law, Review of Law and Economics 5, 33–52.

Levati, M. V., Morone, A., & Fiore, A. (2008), Voluntary contributions with imperfect information: An experimental study, Public Choice 138, 199–216.

Miceli, T.J. (1991), Optimal criminal procedure: Fairness and deterrence, International Review of Law and Economics 11, 3–10.

Newman, J. (1993), Beyond "reasonable doubt", 68 N.Y.U. Law Review 979.

Ostrom, E., Walker, J., & Gardner, R. (1992), Covenants with and without a sword: Self-governance is possible, American Political Science Review 86, 404–417.

Png, I.P.L. (1986), Optimal subsidies and damages in the presence of judicial error, International Review of Law and Economics 6, 101–105.

Polinsky, A.M., & Shavell, S. (1989), Legal errors, litigation, and the incentive to obey the law, Journal of Law, Economics, & Organization, 5, 99–108.

Polinsky, A.M., & Shavell, S. (2000), The economic theory of public enforcement of law, Journal of Economic Literature 38, 45–76.

Polinsky, A.M., & Shavell, S. (2007), The theory of public enforcement of law, in: Polinsky, A.M., & Shavell, S. (eds.), Handbook of Law and Economics, Amsterdam.

Rachlinski, J., & Jourden, F. (2003), The cognitive components of punishment, Cornell Law Review 88, 457–485.

Rizzolli, M., & Saraceno, M. (2009), Better X guilty persons escape than that one innocent suffer, working paper, University of Milan – Bicocca.

Rubinfeld, D., & Sappington, D. (1987), Efficient awards and standards of proof in judicial proceedings, The RAND Journal of Economics 18, 308–315.

Sanchirico, C.W. (1997), The burden of proof in civil litigation: A simple model of mechanism design, International Review of Law and Economics 17, 431–447.

Shavell, S. (2004), Foundations of Economic Analysis of Law, Harvard.

Shaw, M. (2008), International Law, Cambridge.

Tillers, P., & Gottfried, J. (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.

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

United States Supreme Court (1982), Santosky v. Kramer, 455 U.S. 745.

Volokh, A. (1997), N guilty men, University of Pennsylvania Law Review 146, 173–216.

Wistrich, A., Guthrie, C., & Rachlinski, J. (2005), Can judges ignore inadmissible information? The difficulty of deliberately disregarding, University of Pennsylvania Law Review 153, 1251–1345.

Yamagishi, T. (1986), The provision of a sanctioning system as a public good, Journal of Personality and Social Psychology 51, 110–116.

Zelmer, J. (2003), Linear public goods: A meta-analysis, Experimental Economics 6, 299–310.

Zuckerman, A.A.S. (1986), Theories of inference and adjudication: Law, fact or justice? 66 B. U. L. Rev. 487.

Appendix A: Instructions for the P/.5 treatment[19]

General explanations for participants

You are taking part in an economic science experiment. If you read the following explanations carefully, you can earn a rather significant sum of money, depending on the decisions you make. It is therefore very important that you pay attention to the following points.

The instructions you have received from us are intended solely for your private information. *During the experiment, you will not be allowed to communicate with anyone.* Should you have any questions, please direct them directly to us. Not abiding by this rule will lead to exclusion from the experiment and from any payments.

In this experiment, we calculate in taler, rather than in Euro. Your entire income will therefore initially be calculated in taler. The total sum of taler will later be calculated in Euro as follows:

10 taler = 25 Euro cents

The euro you will have accrued plus 2.50 Euro for your participation will be paid to you *in cash* at the end of the experiment.

The experiment is divided into separate periods. It consists of a total of 10 periods. Participants are randomly assigned to groups of four. Each group, thus, has three other members, apart from you. During these 10 periods, the constellation of your group of four will remain unaltered. *You will therefore be in the same group for 10 periods.* Please note that the identification number assigned to you and the other members of the group changes randomly in each period. Therefore, group members cannot be identified as the periods progress.

Each participant will receive from us one installment of 50 taler, with which you will be able to counterbalance potential losses.

---

[19] Differences to the other treatment conditions are marked with footnotes. Screens differed accordingly.

The following pages outline the exact procedure of the experiment.

Information on the exact procedure of the experiment

Each of the ten periods contains two steps.[20]

Step 1[21]

At the beginning of each period, each participant is allotted *20 taler,* which we shall henceforth refer to as his *endowment.* The player's job is now to make a decision with regard to using his endowment. You have to decide how many of the 20 taler you wish to pay into a *project* and how many you wish to keep for yourself. The consequences of your decision are explained in greater detail below.

At the beginning you will see the following input screen of Step 1:

The input screen of step 1


```
┌─Period──────────────────────────────────────────────────────────────────────────────────┐
│              1                                              Remaining time [sec]:  118     │
├───────────────────────────────────────────────────────────────────────────────────────────┤
│                                                                                           │
│  In this period you are group member 1                                                    │
│                                                                                           │
│  Step 1                                                                                    │
│                                                                                           │
│  You receive in this period an endowment of 20 taler.                                      │
│                                                                                           │
│  Out of these 20 taler you pay …                                                          │
│                                                                                           │
│                            … into the project:  [      ]                                  │
│                                                                                           │
│                                                                        [  OK  ]           │
└───────────────────────────────────────────────────────────────────────────────────────────┘
```

In the left upper corner of the screen you will find the *period number*. In the right upper corner you will find the *remaining time* for your decision *in seconds*.

---

[20] Sentence is missing in N/.5.
[21] Missing in N/.5.

Your *endowment* is, thus, *20 taler in each period*. You make a decision on your project contribution by typing any one whole number between 0 and 20 into the appropriate field on your screen. This field can be accessed using the mouse. As soon as you have determined your contribution, you have also decided on how many taler to keep for yourself, i.e., *20 – your contribution*. Once you have typed in your contribution, please click on *OK*, again using the mouse. Once you have done this, your decision for this period is irreversible.

Your income therefore consists of two parts, namely:

(1) the taler you have kept for yourself
(2) the *"income gained from the project".* Your income from the project is calculated as follows:

*Income from the project = .4 * total sum of all contributions to the project*

Your *income in Step 1*[22] of each period equals:

Income in Step 1[23] *=*

20 (endowment)
– Your contribution to the project
+.4(total sum of contributions to the project)*

The total income in Step 1,[24] in taler, is calculated using the same formula for each member of the group. If, for example, the sum of the contributions from all group members adds up to 60 taler, you and all other members each receive a project income of .4x 60 = 24 taler. If the group members have contributed a total of 9 taler to the project, you and all other members each receive an income of .4x9 = 3.6 taler from the project.

For each taler you keep for yourself, you earn an income of 1 taler. If, on the other hand, you contribute one taler from your endowment to your group's project instead, the sum of the contributions to the project increases by one taler and your income from the project increases by .4x1 = .4 taler. However, the income of each individual group member also increases by .4 taler, so that the group's total income increases by .4x4 = 1.6 taler. The other group members thereby also profit from your contributions to the project. In turn, you

---

[22] "Income" in N/.5.
[23] "Income" in N/.5.
[24] "Income" in N/.5.

25

profit from other members' contributions to the project. For each taler contributed to the project by another group member, you earn .4x1 = .4 taler.

Step 2[25]

In Step 2, you can decrease each individual group member's income by giving points, or leave as it is. For this purpose, you receive an additional endowment of 10 taler which you may use to buy points. These taler form the additional endowment; any taler you do not use for acquisition of points are kept and added to your period income at the end of step 2. All other group members are allowed to decrease your income, too, if they so wish.

You may assign points in the input screen of step 2 which shows, along the number of periods and the remaining time, for each group member a signal about the contribution to the project. Your contribution will be shown in the row "You".

The input screen of Step 2

Period		
1		Remaining time [sec]: 107

Step 2

Group member	Signal	Points
You	XXX	
Group member 2	YYY	
Group member 3	YXY	
Group member 4	YYX	

OK

Please notice that the signals for the three other group members only have a 50% probability of equaling the actual contribution. This means that the signaled contribution to the project for each of the other group members equals their actual contribution in 5 out of 10 cases on average. There is a 50% probability that you will see a random, different number which does not correspond to the particular group member's contribution. In this

---

[25] Unless indicated, the section "Step 2" is missing in N/.5.

case, there is an equal probability that any other number than the actual contribution will appear.[26]

You now have to decide how many points you wish to distribute to each group member. You must enter a number at this stage. If you do not wish to alter a certain group member's income, please enter 0. If you want to assign points, you have to choose a number greater than 0. You can operate within the fields by using the tab key (→|) or the mouse. You may still change your decision as long as you have not yet clicked on *OK*.

When distributing points, you incur costs in taler which depend on the number of points you distribute to the individual players. The more points you distribute, the higher your costs are. The following formula shows the connection between the points distributed and the costs of such distribution:

Costs for assigned points = sum of points (in taler)

Each distributed point costs you 1 taler. Therefore, you can distribute a maximum of 10 points. For example, if you have allocated 2 points to one member, your cost is 2 taler; if, in addition, you distribute 7 points to another group member, your cost is 7 taler; if you give the final group member 0 points, you have no costs. The *total cost* to you is therefore 9 taler (2+7+0). The remaining rest of the additional endowment for buying points, in this example 1 taler, is kept and added at the end of step 2 to your period income.

If you choose 0 points for a certain group member, you do not alter this group member's income. If you allocate 1 point (choosing 1) to a group member, you *decrease* this particular group member's taler income from Step 1 by *3 taler*. If you allocate 2 points to a group member (choosing 2), you decrease his income by 6 taler etc. *Each point allocated by you to a particular group member reduces the group member's taler income from Step 1 by 3 taler.*

The *overall* reduction in a group member's income from Step 1 depends on the total number of points received. If, for instance, one member receives *a total of 3 points* from all other members, the income in Step 1 is reduced by *9 taler*. If a member receives a total of *4 points*, the income in Step 1 is reduced by *12 taler.* A person who receives points will be informed about the income reduction at the end of each period, without detailed information on the group member who distributed that point.

---

[26] In P/.9, "5" is replaced by "9" in the paragraph. In P/1, the paragraph is replaced by "The other signals correspond to the contributions of all other three group members."

For your total income at the end of step 2, it follows that:

Total taler income at the end of step 2 = Period income =

Income after step 1

+ 10 (additional endowment for buying points)

− 3*(sum of received points)

− cost of points you distribute

Please note that your total taler income at the end of step 2 can be negative if the costs for distributed points exceed the income after step 1 minus the reduction of income due to received points.

The income screen at the end of Step 2

Period	
1	Remaining time [sec]: 115

Period income

Taler you have kept for yourself	xxx
Your income from the project	yyy
Your income in Step 1	XXY
Additional endowment for buying points	10
Costs for points you distributed	-yxy
Points you received	xyx
Resulting reduction in your income (in taler)	-yyx
Your period income	YXX

Continue

Once[27] all members of the group have made their decisions, you will be informed about your period income in the income screen at the end of Step 2.[28] Here, you see how many taler you kept for yourself, your income from the project, and the resulting income in Step 1.[29] In addition, you are informed about the costs for points you distributed, the number of points you received, as well as the resulting reduction in income. [30] Finally, you will see your period income.

---

[27] In N/.5, the following paragraph, screen, and formula are included in the instruction.
[28] In N/.5, "…in the income screen."
[29] In N/.5, "…and the resulting income."
[30] In N/.5, the sentence is replaced by "In addition, you receive an additional endowment of 10 taler that will be added to your income."

Your total income at the end of the experiment equals the sum of all period incomes:

Total income (in taler)

= Total sum of period incomes
(If the sum of period incomes is negative, your income is 0 taler.)

Do you have any further questions?

Appendix B: Control questions[31]

Please answer all questions. There are no consequences for you due to wrong answers. If you have any questions, please contact us.

1. Each group member is endowed with 20 taler. None (including you) contributes anything in step 1.
   - What is your income in step 1?
   - What is the income of each of the other group members in step 1?
2. Each group member is endowed with 20 taler. Each group member (including you) contributes 20 taler to the project in step 1.
   - What is your income in step 1?
   - What is the income of each of the other group members in step 1?
3. Each group member is endowed with 20 taler. The other three group members contribute a total of 30 taler to the project in step 1.
   - What is your income in step 1 if you contribute – in addition to the 30 taler – 0 taler to the project?
   - What is your income in step 1 if you contribute – in addition to the 30 taler – 15 taler to the project?
4. Each group member is endowed with 20 taler. You contribute 8 taler to the project.
   - What is your income in step 1 if the others group members contribute – in addition to your 8 taler – in total 7 taler to the project?
   - What is your income in step 1 if the others group members contribute – in addition to your 8 taler – in total 22 taler to the project?
5. In step 2, you distribute points to each of the three other group members: 3, 5, and 0 points. What are the total costs for the distribution of those points?

---

[31] Questions 5 to 9 are missing in N/.5.

6. What are the total costs if you distribute to all group members 0 points?

7. What is the reduction of income in step 1 if you receive in total 0 points?

8. What is the reduction of income in step 1 if you receive in total 4 points?

9. What is the reduction of income in step 1 if you receive in total 15 points?