

Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement

Andreas Fuster and Stephan Meier

Abstract:

Monetary incentives are often considered as a way to foster contributions to public goods in society and firms. This paper investigates experimentally the effect of monetary incentives in the presence of a norm enforcement mechanism. Norm enforcement through peer punishment has been shown to be effective in raising contributions by itself. We test whether and how monetary incentives interact with punishment and how this in turn affects contributions. Our main findings are that free riders are punished less harshly in the treatment with incentives, and as a consequence, average contributions to the public good are no higher than without incentives. This finding ties to and extends previous research on settings in which monetary incentives may fail to have the desired effect.

JEL Classifications: C72, C92, D23, H41

Keywords: public goods, experimental economics, norm enforcement, hidden costs of incentives

Andreas Fuster is a graduate student in the economics department at Harvard University and a research associate in the research department of the Federal Reserve Bank of Boston. Stephan Meier is an assistant professor at the Graduate School of Business of Columbia University and a visiting scholar in the Research Center of Behavioral Economics and Decisionmaking at the Federal Reserve Bank of Boston. Their email addresses are afuster@fas.harvard.edu and sm3087@columbia.edu, respectively.

This paper, which may be revised, is available on the website of the Federal Reserve Bank of Boston at <http://www.bos.frb.org/economic/wp/index.htm>.

The views expressed in this paper are solely those of the authors and not necessarily those of the Federal Reserve Bank of Boston or the Federal Reserve System.

We are grateful to Luke Coffman, Armin Falk, Ernst Fehr, Simon Gächter, Lorenz Goette, and seminar audiences at Harvard University and the University of Zurich for helpful comments and discussions, and to Benjamin Levinger for help in conducting the experiments.

This version: March 3, 2009

1 Introduction

Prosocial behavior, such as making private contributions to public goods, is crucial for the proper functioning of society and the efficiency of organizations. Various activities performed by humans, from hunting to holding a potluck party, require discretionary contributions of the group members to be successful. Many common pool resources become exhausted if individuals do not refrain from consuming the privately optimal but socially suboptimal amount. Similarly, the success of organizations depends on members' willingness to take unselfish, efficiency-enhancing actions, because it is difficult to fully control behavior through contracts. However, the private incentives to free ride in such situations often make it difficult to sustain high levels of prosocial behavior. Recent research shows that peer punishment or norm enforcement, that is, the willingness of people to incur a cost to punish a free rider, can potentially explain the maintenance of high levels of cooperation (Fehr and Gächter, 2000, 2002).¹ For instance, a norm enforcement mechanism can successfully prevent the exploitation of common pool resources (Ostrom, 1990), while within firms, social pressure and mutual monitoring can play an important role in inducing effort and thereby contribute crucially to an organization's efficiency (Mas and Moretti, 2009). However, if a norm enforcement mechanism is not working properly, social dilemmas will not be solved, and indiscriminate punishment can lead to large welfare losses (Herrmann et al., 2008).

This paper investigates in a laboratory experiment how private monetary incentives for contributing to a public good affect norm enforcement. Monetary incentives are widely used by policymakers and managers to foster prosocial behavior. For example, recycling or the use of environmentally friendly technologies (such as hybrid cars) is often subsidized. Within organizations, teamwork, contributions to the work

¹In situations where the punisher knows that he will never interact with the free rider again, this behavior is referred to as "altruistic punishment".

atmosphere, and extra-role behavior are often incentivized with private monetary incentives such as bonuses, awards, or promotions. Offering private incentives to behave prosocially reduces the price of prosocial behavior for contributors, and this has been found to be effective in raising contributions in a number of situations, as standard economics would predict.² However, a growing body of evidence in psychology (for a survey, see Deci et al., 1999) and economics (for surveys, see Frey and Jegen, 2001; Bowles, 2008) finds that monetary incentives can crowd out individuals' willingness to behave prosocially, leading to a *direct* detrimental effect of monetary incentives. Our paper instead focuses on the question of whether private monetary incentives, such as bonuses for extra-role behavior in a firm, can have an *indirect* effect on the level of prosocial behavior by affecting the functioning of a norm enforcement mechanism.

How private monetary incentives affect overall prosocial behavior in a setting where norm enforcement is important depends on how the incentives affect two crucial factors of norm enforcement: the propensity of prosocial individuals to inflict costly punishment on free riders, and the reaction of free riders to such punishment. If the incentives do not influence those factors, the effects of the incentives and of norm enforcement will be additive and a combination of the two should be most successful in fostering high levels of prosocial behavior. However, as we argue below, monetary incentives can also dampen the effectiveness of the norm enforcement mechanism, leading to less punishment and as a consequence lower contribution rates — even with an additional incentive in place.³

How could monetary incentives *negatively* influence the two factors of a successful norm enforcement mechanism? First, norm enforcement mechanisms depend on the

²For example, tax deductions have been shown to increase charitable giving (Auten et al., 2002) and in experimental studies, people reacted consistently to changes in prices (Andreoni and Miller, 2002).

³Obviously, monetary incentives may also improve the effectiveness of the punishment mechanism, as we will discuss in detail when developing the behavioral hypotheses.

willingness of some individuals to punish free riders.⁴ It seems that high contributors are motivated to punish free riders because it allows them to vent their anger and express disapproval (Bosman and van Winden, 2002; Fehr and Gächter, 2002; Hopfensitz and Reuben, 2009), and that they derive satisfaction from the act of punishing norm violations (de Quervain et al., 2004). An extra incentive for a prosocial individual can potentially mitigate his anger, as the advantage of free riding is reduced by the incentive reward. For example, in the absence of private incentives for exerting effort that benefits the firm (and therefore, directly or indirectly, all its employees), hard-working employees may feel angry when observing free riders who receive the same salary, leading to norm enforcement efforts. However, if employees are rewarded for hard work and extra-role behavior, their willingness to punish may be reduced, as they receive something that free riders do not.

Second, a successful punishment mechanism also relies on free riders adapting their behavior in response to punishment. For instance, employees who do not contribute their “fair share” to the public goods in a firm may feel compelled to increase their contributions after being sanctioned by their peers. But in the presence of private incentives for contributors, free riders may not feel guilty for not contributing, and perceive punishment as unjustified, as they already forgo the additional incentive. As a result, free riders may not increase their contributions after punishment as much as when no private incentives are in place. Thus, if private monetary incentives dampen one or both of the factors that together sustain prosocial behavior over time, the norm enforcement mechanism may not be as effective as when no private incentives are present, and this may lead to lower overall prosocial behavior.

The results from our experiment show that an exogenous introduction of a relatively substantial monetary contribution incentive can indeed negatively affect both factors

⁴Throughout the paper, we use the term “free rider” to refer to individuals whose contributions to the public good are below average.

of the norm enforcement mechanism. First, offering salient monetary incentives proportional to contributions leads to less severe punishment of free riders. In the setting with monetary incentives, deviations from the group average are punished less harshly than in a setting where no monetary incentives are offered. Second, punishment has less of an influence on free riders' behavior when monetary incentives for contributors are in place. For each punishment point received, free riders increase their subsequent contribution by less when monetary incentives are offered than when no incentives are offered. As a result of the negative reactions of both factors to incentives, prosocial behavior is not increased by the non-trivial incentive we provide — even though the incentive increases contributions substantially in the absence of norm enforcement.

These findings indicate that policymakers and managers should be careful in using private incentives to foster prosocial behavior in settings where norm enforcement and social pressure are important. We therefore contribute to the discussion about possible “hidden costs” of incentives (Lepper and Greene, 1978). A number of studies show direct negative effects of incentives on individuals' prosocial behavior (for example, Frey and Oberholzer-Gee, 1997; Gneezy and Rustichini, 2000a,b; Mellström and Johannesson, 2008), and a variety of potential explanations for these negative effects have been suggested.⁵ The results from our experiment show that incentives may not only affect prosocial behavior *directly* but also *indirectly*, through their negative influence on the effectiveness of norm enforcement. Monetary incentives can therefore have a strong positive effect in settings where there is no norm enforcement, while the effect of incentives on prosocial behavior is weaker or even negative when norm enforcement

⁵Extrinsic motivations might interact with intrinsic motivations (for example, Deci, 1975; Frey, 1997); extrinsic incentives might destroy trust in a principal-agent relationship (for example, Fehr and Falk, 2002; Fehr and List, 2004; Falk and Kosfeld, 2006); the introduction of extrinsic motives might shift an individual's decision frame from a social frame to a monetary frame (for example, Gneezy and Rustichini, 2000a; Heyman and Ariely, 2004), or monetary incentives might influence individuals' image motivation (for example, Benabou and Tirole, 2006; Ellingsen and Johannesson, 2008; Ariely et al., 2009).

is important and, by itself, powerful. Having said that, it is not clear in our setting that the presence of incentives does not increase welfare, even though contributions are not increased — after all, if a similar level of prosocial behavior can be achieved with less (socially wasteful) norm enforcement, this is welfare-enhancing. On the other hand, while norm enforcement is costly in the short run, in the long run it is mainly the threat of punishment that maintains high levels of contributions, while providing incentives continues to be costly. Therefore, whether the introduction of private incentives for prosocial behavior is desirable for a policymaker or manager will depend on the weights she assigns to contributions versus norm enforcement costs, on the cost of providing incentives, and on the horizon over which the policy is considered.

The remainder of the paper is structured as follows: In the next section, we describe our experimental design and its two treatments, the Baseline Treatment and the Incentive Treatment. Section 3 discusses our behavioral hypotheses, namely the various ways in which the presence of incentives could influence the different aspects of the norm enforcement mechanism. Section 4 then presents the results from our experiment, and Section 5 discusses these further. Finally, we briefly conclude and suggest topics for further research.

2 Experimental Design

Our experiment consists of a linear public good game (also known as “Voluntary Contribution Mechanism”) with four treatment conditions (see Table 1). In the *Baseline Treatment (BT)* subjects participate in two six-period public good games with and without punishment opportunity. In the *Incentive Treatment (IT)*, a private monetary incentive is added to the BT.

2.1 The Baseline Treatment(BT)

In the *Baseline Treatment*, participants first play six periods of a public good game in fixed groups of four. In each period, each group member $i \in \{1, 2, 3, 4\}$ receives an endowment of 20 Experimental Currency Units (ECU) and can contribute an integer g_i ($0 \leq g_i \leq 20$) to a public good (referred to as a “group project”). All group members decide simultaneously on their g_i in a period. The monetary payoff of each individual i from the group project in a period is given by

$$\pi_i^1 = 20 - g_i + a \sum_{j=1}^4 g_j, \quad (1)$$

where a is the marginal per capita return from a contribution to the public good. In this experiment, as in Fehr and Gächter (2000, 2002) and many subsequent papers in this literature, a is set to equal 0.4. Hence, the private cost to an individual of contributing 1 ECU to the public good is 0.6 ECU, while the total benefit to his fellow group members is 1.2 ECU. This means that not contributing at all ($g_i = 0$) is the dominant action for each group member i in the stage game, while the total group payoff ($\sum_{i=1}^4 \pi_i^1$) is maximized if all group members contribute their full endowment ($g_i = 20$).

After six periods without punishment, participants are re-matched into new groups and play another six-period public good game with the same parameter values as in the first six periods, but with a peer punishment mechanism (implemented as in Fehr and Gächter, 2002).⁶ In each period, participants now receive an additional endowment of 10 ECU.⁷ After participants make their contribution decision g_i , they are informed about the contribution of each other group member $j \neq i$, and are allowed to assign

⁶Participants were only informed about the second six-period game once the first game was over.

⁷This was done to reduce the likelihood that a subject would refrain from punishment in order to avoid the risk of a negative payoff in a period.

punishment points, p_{ij} ($0 \leq p_{ij} \leq 10$), to the other group members. The punishment points (neutrally labeled deduction points in the instructions) are costly to both the sender and the receiver. Each punishment point costs 1 ECU to the sender and 3 ECU to the receiver. However, the payoff-effective punishment costs imposed by the other group members on subject i , C_i , cannot exceed the first-stage payoff, π_i^1 . C_i is therefore given by $C_i = \min(3 \sum_{j \neq i} p_{ji}, \pi_i^1)$. The overall payoff of subject i in a period of the public good game with punishment is then given by

$$\pi_i = 10 + \pi_i^1 - C_i - \sum_{j \neq i} p_{ij}. \quad (2)$$

2.2 The Incentive Treatment (IT)

The *Incentive Treatment* is identical to the *Baseline Treatment*, except that the subjects are now given a private monetary incentive to contribute to the public good. To make the private incentive salient, participants receive a (virtual) “lottery ticket” for each ECU that they contribute to the public good. Each lottery ticket gives a 1 percent chance of winning an additional 20 ECU at the end of the experiment, so it has an expected value of 0.2 ECU.⁸ For example, if a participant contributes 10 ECU to the public good, he or she will, in expectation, win 2 ECU. Thus, the expected private monetary payoff in the IT is increased by $0.2g_i$ compared with the payoffs in the BT. This incentive is in place for both parts of the treatment, with and without punishment.

The monetary incentive is non-trivial, as, in expectation, it is equivalent to a reduction from 0.6 ECU to 0.4 ECU in the private cost of contributing an ECU to the public good. It is important to note, however, that this incentive does not alter the benefits that the other members of i 's group derive from i 's contribution, and that from

⁸The lotteries were conducted so that each subject's tickets were “additive,” not independent, within each six-period block of the treatment. This means that having x tickets gives a one-time x percent chance to win 20 ECU, rather than x independent 1 percent chances of winning.

a purely monetary perspective, it is still a dominant strategy in the stage game for each subject to not contribute anything to the public good, unless a subject is extremely risk-loving.⁹

2.3 Procedures

The experiments were conducted at the Computer Lab for Experimental Research at Harvard University using the software z-Tree (Fischbacher, 2007). 60 subjects, the vast majority of them undergraduate students, participated in the BT and 76 in the IT. Participants received detailed written instructions with a number of control questions, and the experiment started only after all participants answered all the questions correctly.¹⁰ At the end of a session, one period of each six-period game was randomly chosen to determine the final payoff (consisting of the monetary payoff, and in the IT, the number of lottery tickets), and (in the IT) the two lotteries were played out. Then, participants were paid their total earnings from the games, converted at a rate of 1 ECU = US\$ 0.25, and a show-up fee of US\$ 10.00. Average earnings for the experiment, which lasted approximately 80 minutes, were US\$ 24.40.

Using one randomly chosen period for the final payoff has been shown not to affect behavior as compared with paying subjects for each period (Laury, 2005). Paying only one period per game allows us to increase the stake for each decision, and in particular, to offer a non-trivial prize in the lottery in the IT (20 ECU = US\$ 5). In turn, having higher stakes may change behavior as compared with a situation with lower stakes. Most other papers in the literature pay subjects for every period, but at much lower conversion rates (for instance, Herrmann et al., 2008, convert at the rate

⁹We believe it is very unlikely that a subject would be willing to give up $0.6x$ ECU for an x percent chance of winning 20 ECU, which is what would be needed to make contributing (at least) x the dominant action.

¹⁰The instructions distributed to the subjects can be found in the appendix.

1 ECU = US\$ 0.03). This may explain why we find somewhat lower contribution and punishment levels than most other papers in this literature.

3 Behavioral Hypotheses

The main goal of this paper is to investigate how the presence of private incentives to contribute to the public good affects the functioning of a norm enforcement mechanism (peer punishment) and the resulting level of public good contributions. Previous research indicates that the ability of a peer punishment mechanism to sustain or increase public good contributions depends largely on two factors: (1) how harshly subjects who contribute less than average (“free riders”) are punished, and (2) how those free riders adapt their contributions afterwards.¹¹ If private incentives do not interact with either of these factors, we should expect that the IT yields higher contributions than the BT, with and without the punishment mechanism. This is because the private cost of contributing is reduced, which should bring contributions to a higher level (a standard “price effect”).¹² However, as we discuss below, the presence of private incentives may influence punishment behavior and the reaction of free riders, and as a result, these interactions could yield higher or lower overall contribution levels than would be expected from the price effect alone.

¹¹The extent of “antisocial punishment” (meaning the punishment of above-average contributors) is also crucial, as shown by Herrmann et al. (2008). However, because we do not expect (or find) much antisocial punishment in our American subject pool, we do not focus on it in our discussion.

¹²Of course, this already assumes that subjects do not act in accordance with standard game theory, which predicts that nobody ever contributes (or punishes). Various papers have looked at how contributions are affected by changes in the marginal per capita return, a , in public good experiments without punishment, and they generally find a fairly strong and significant price effect (see the survey in Ledyard, 1995). A recent paper by Carpenter et al. (2009), which varies a as well as group size in a “stranger” setting with punishment, finds that an MPCR of 0.75 rather than 0.3 leads to higher mean contributions when the group size is eight but not when it is four. As discussed earlier, we do not change a across treatments, but rather give a personal incentive akin to a rebate in the IT.

3.1 The Punishment of Free Riders

A growing body of research has investigated the driving forces behind punishment behavior. One of the main findings is that negative emotions towards free riders are an important motivation, or perhaps even the main motivation, for punishment. High contributors punish free riders to vent their anger and express their disapproval (Bosman and van Winden, 2002; Fehr and Gächter, 2002; Hopfensitz and Reuben, 2009), and derive satisfaction from doing so (de Quervain et al., 2004).¹³

Depending on exactly what determines the strength of negative emotions towards free riders, monetary incentives for contributing to the public good may either increase, have no effect on, or decrease punishment.¹⁴ To illustrate the three different possible effects of incentives on punishment, it may be helpful to consider the following simple one-shot, two-person social dilemma. First-stage payoffs are given by $\pi_i^1 = E - (1 - r)g_i + a \cdot (g_i + g_j)$, where E is the subjects' endowment, g_i the contribution of subject i , r the private reward a subject receives per unit of contribution, and a the marginal per capita return (MPCR) of the public good.¹⁵ Assume that $g_i > g_j$ and that we are interested in i 's punishment decision. What determines i 's anger towards j ?

¹³On the other hand, pure strategic reasoning (namely, punishing free riders in order to lead them to increase their future contributions) seems to be less important for punishment. High contributors punish free riders even in a pure stranger design (where subjects are certain never to be in a group with the same other subjects again) (Fehr and Gächter, 2002; Egas and Riedl, 2008) or when sanctions are revealed only at the end of the experimental session (Vyrastekova et al., 2008). Nor does reducing payoff inequalities seem to be main motive behind punishment, as subjects punish even when the punishment has no consequences for inequality (Falk et al., 2005; Masclet and Villeval, 2008; Egas and Riedl, 2008). In other settings, it has been shown that others' perceived intentions, and not just the consequences of their actions, matter for people's punishment behavior (Brandts and Charness, 2003).

¹⁴Relatively little is known about what determines the strength of emotional reactions in such settings. It is known that the strength of anger is strongly related to people's expectations — opportunistic acts that are more unexpected cause stronger anger and (if available) harsher punishment (see, for instance, Hopfensitz and Reuben, 2009, and references therein). The following discussion can therefore be seen as an exploration of what determines expectations.

¹⁵ a and r are such that $(1 - r)/2 < a < 1 - r$, which makes this a social dilemma, as it is in each subject's private monetary interest to choose $g_i = 0$, while total payoff $\pi_i^1 + \pi_j^1$ is maximized if $g_i = g_j = E$.

Case 1: The first possibility is that high contributors become angry at people they see as selfish, and want to punish them for their selfish traits, à la Levine (1998). In terms of our social dilemma, we could define j 's "selfishness" s_j as the 'benefit withheld from the other player' divided by the unit cost of providing this benefit, or

$$s_j = \frac{a(E - g_j)}{1 - r}, \quad (3)$$

and we can then imagine that i 's punishment of j increases in j 's selfishness as compared to i 's,

$$s_j - s_i = \frac{a(g_i - g_j)}{1 - r}. \quad (4)$$

Clearly, this expression increases in r . Intuitively, not contributing (or contributing little) is more selfish the lower the personal cost of contributing. Thus, this would predict that there should be more punishment in the IT because not contributing is a more selfish action there.

Case 2: Alternatively, high contributors may punish based on the harm that the low contributors' actions impose on them, without taking into account the cost of contributing. In terms of our example, this would mean that i punishes j solely based on the numerator of the previous expression, $a(g_i - g_j)$, which is independent of r . For our experiment, where the MPCR a is the same in both treatments, this would predict equal strength of punishment in both treatments.

Case 3: Finally, it may be that inequality of outcomes is what triggers negative emotions towards low contributors and therefore punishment. In our example, we have:

$$\pi_j^1 - \pi_i^1 = (1 - r)(g_i - g_j), \quad (5)$$

which is decreasing in r , meaning that the payoff inequality for a given contribution

difference is smaller if contributing is less costly. In our experiment, this would predict that punishment of free riders is less harsh in the IT than in the BT, as high contributors receive a private reward in the former but not in the latter. In light of the earlier discussion of the motives of punishment, it is important to note here that even though Falk et al. (2005) and others have shown that reducing the payoff inequality between the punisher and the target is not the main *goal* of punishment, it may still be the case that inequality is the *cause* of the negative emotions that lead to punishment.¹⁶ This would be consistent with the findings of Dawes et al. (2007), who look at punishment in a setting where first-stage payoffs are randomly generated (in fact, drawn from the distribution of first-stage payoffs in Fehr and Gächter, 2002) rather than determined by contribution decisions. Dawes et al. find that high earners are still punished substantially, even though their high earnings are not a result of free riding, and that in a hypothetical scenario, their subjects express negative emotions (annoyance and anger) towards high earners, the more strongly the higher the inequality.

In summary, under different assumptions of what drives punishment, we would predict different effects of private contribution incentives on the strength of punishment. Our results will therefore not only show whether private incentives interact with punishment behavior but also give an indication of which assumption regarding the causes of negative emotions towards free riders is most reasonable (at least in the public good context used in our experiment).

¹⁶This is noted, for instance, by Fehr and Gächter in their response to Fowler et al. (2004), who point out that “egalitarian motives,” as opposed to negative emotions towards free riders, may be responsible for the punishment observed in Fehr and Gächter (2002): “Fowler et al. contrast their egalitarianism hypothesis with our view that negative emotions against free riders drive punishment. However, the two views are not necessarily incompatible: egalitarian sentiments may be the basis behind cooperators’ negative emotions because free riding causes considerable inequalities.” Fehr and Gächter go on to point out that egalitarian motives cannot explain the results in Falk et al. (2005), as mentioned earlier.

3.2 The Reaction of Free Riders

The success of a peer punishment mechanism in increasing contributions depends not only on sufficiently harsh punishment of free riders, but also on their reaction to the punishment. It has been observed in numerous experiments that free riders who are punished in a period increase their contribution in the subsequent period. One reason for doing so is surely to avoid the material costs from further punishment. Yet, this may not be the only motivation: a free rider may also increase his contributions because he feels bad for not adhering to the contribution norm of his group, and this feeling may be enhanced if he is punished by the other group members, who thereby clearly signal their disapproval of the free rider's action. Bowles and Gintis (2005) refer to these two respective feelings as guilt and shame, and argue that these emotions play a significant role in increasing the contributions of free riders.

Direct evidence for the possible importance of guilt and shame in a public good experiment is provided by Masclet et al. (2003), who investigate the effect of nonmonetary sanctions. In their experiment, subjects can express their disapproval about the actions of another group member by assigning "disapproval points" without any monetary costs to either the sender or the receiver of these points. The results show that free riders who receive more disapproval points increase their contributions by more in the next period, consistent with the "shame" hypothesis. This is the case even in a "stranger" setting, where subjects are re-matched into new groups in every period. Another finding is that free riders who are furthest below the average contribution in the previous period increase their subsequent contributions the most for a given level of punishment, a finding that is consistent with the "guilt" hypothesis.

Hopfensitz and Reuben (2009) provide further evidence by looking at a one-shot, two-person trust game with punishment and possible counter-punishment. Hopfensitz and Reuben measure various emotions (such as anger, guilt, shame, surprise) of the

players directly through questionnaires right after players observe their partner’s action, but before they choose their own action. They find that among the second movers who are punished for defecting (meaning that they returned little of the entrusted money), those who reported feeling guilty were less likely to retaliate than those who did not feel guilty, and they returned more money when playing the game a second time (against a different first mover). Furthermore, the intensity of guilt expressed by second movers seems independent of whether they were punished. These results support the claim that “prosocial emotions” such as guilt and shame are crucial for the effectiveness of a punishment institution.

For our experiment, this means that the effect of private incentives on contributions may depend on how these incentives influence the extent to which free riders feel guilty or ashamed for not contributing as much as their peers. In ways similar to the ones discussed in the previous section, it is conceivable that the presence of private incentives enhances or reduces the strength of these emotions or leaves them unaffected. Thus, if private incentives increase the perceived selfishness of free riders and as a consequence their guilt and shame as in *Case 1*, then the reaction of free riders to a given level of punishment will be stronger. However, if inequality drives not only the anger of the high contributors but also the guilt and shame of free riders (as in *Case 3*), then the presence of incentives will lower free riders’ guilt and shame and will dampen their reaction to punishment.¹⁷

The overall effect of private incentives on the contribution level in the public good

¹⁷A recent paper by Reuben and Riedl (2009) considers a different twist on the public good game with punishment and contains findings consistent with that last possibility. Reuben and Riedl experimentally investigate contributions and sanctioning in “privileged groups,” meaning that for one group member it is privately optimal to contribute his full endowment to the public good, because the benefit he derives from it is sufficiently high. They find that such privileged groups obtain significantly higher contributions than “normal” groups when no peer punishment is possible, but that this is reversed with peer punishment. The main reason for this seems to be that the “low-benefit” subjects in privileged groups are less willing to increase their subsequent contributions in response to being punished than the “low-benefit” subjects in “normal” groups (where all group members are “low-benefit”).

game will then depend on the combination of three factors: (1) The “price effect” of the incentive (it becomes cheaper to contribute); (2) the intensity of the punishment inflicted on free riders, and (3) the change in free riders’ contribution behavior over time. If (2) and (3) are unaffected by the presence of incentives, we expect contribution levels to go up, in magnitude comparable to the incentive effect in a setting without peer punishment. If punishment becomes harsher and/or free riders react more strongly to it (in the sense of increasing their subsequent contributions by more), then the private incentives will have a larger positive effect on contribution levels than in a version of the game without punishment — in other words, incentives and peer punishment will be *complements*. However, if instead the presence of incentives leads to weaker punishment of free riders and/or if (punished) free riders are less prone to increase their contributions over time, the positive effect of incentives will be diminished, and if the effects are sufficiently strong, contributions may even be lower than without incentives.¹⁸ In that case, one could say that incentives and peer punishment are *substitutes*.

4 Results

4.1 Effect of Incentives on Contributions

RESULT 1: *Without the punishment mechanism, contributions are significantly higher in the IT than in the BT, while with the punishment mechanism, there is no significant difference between the two treatments.*

Support for Result 1 is presented in Figure 1. Figure 1 shows that in the six periods

¹⁸Of course, it is possible that punishment becomes stronger while the reaction of free riders becomes weaker, or vice-versa. In such a case, the total effect on contributions is ambiguous.

without the punishment mechanism (the left half of the graph), the presence of private incentives leads to significantly higher contributions. On average, contributions are 10.3 ECU with private incentives and only 6.0 ECU without incentives, and a non-parametric Mann-Whitney test rejects the null hypothesis of equal distributions of group average contributions ($z = -2.672, p < 0.01$).

However, in the periods with the punishment mechanism, contributions are not significantly higher when a monetary incentive is added. Average contributions over all periods are 14.2 ECU in the BT and 15.0 ECU in the IT, and a Mann-Whitney test does not reject the null hypothesis of equal distributions ($z = -0.173, p = 0.86$). As can be seen from the right half of Figure 1, contributions are higher in the IT for only the first two periods but are never statistically significantly so at the 95 percent level. If we look only at the last four periods, average contributions are actually slightly higher in the BT (15.4 ECU versus 15.2 ECU).

To gain a better understanding of the differences in contribution behavior between the two treatments when the punishment mechanism is in place, it is useful to compare the proportion of groups that reach the socially efficient outcome of everybody contributing 20 ECU in a period. It may be the case that average contributions in the IT are not significantly higher than in the baseline because of a “ceiling effect,” that is, that participants are unable to increase contributions beyond 20 even though they might be willing to do so if they could. In fact, however, the proportion of groups who reach the socially efficient outcome is *lower* in the IT than in the BT from period 2 onwards (see Figure 2).¹⁹ Thus, the result that private incentives do not increase contributions when a punishment mechanism is in place is not due to a ceiling effect.

To sum up, these results show that incentives have a significant positive effect on contributions when no peer punishment is possible. Thus, our “lottery ticket” incen-

¹⁹The proportions are not quite significantly different at conventional levels, though; the p-value from a one-sided Fisher’s exact test is 0.11 for period 6 and higher for the other periods.

tives “work.” However, with peer punishment, this is no longer the case; contribution levels are indistinguishable across the two treatments, and a somewhat higher proportion of groups manage to reach the efficient outcome in the BT than in the IT. The next two sub-sections compare punishment and subjects’ reaction to it across the two treatments.

4.2 Effect of Incentives on Peer Punishment

RESULT 2: *Group members who contribute less than average (“free riders”) are punished significantly less severely in the IT than in the BT.*

Figure 3 shows the average number of punishment points a subject received as a function of the deviation of the subject’s contribution from the average contribution of the other three group members. In both treatments, free riders are punished, and the more strongly so the further below the average their contributions are. However, this increase in the severity of punishment is much less pronounced in the IT than in the BT. On average, a free rider receives 4.34 punishment points in the BT but only 2.59 points in the IT, and the difference is significant in a Mann-Whitney test ($z = 2.935, p < 0.01$). It is the severity of punishment of free riders that is different across treatments, not the frequency: in the BT, free riders are punished in 75.8 percent of cases while in the IT, they are punished in 71.2 percent of cases, and the difference is not statistically significant ($p = 0.26$, one-sided Fisher’s exact test).

Meanwhile, the number of punishment points received by subjects who contribute more than average does not differ significantly across the two treatments (Mann-Whitney: $z = 0.241, p = 0.8$).²⁰

²⁰The frequency of punishment of such subjects, which is 15.6 percent in the BT and 14.0 percent in the IT, is also statistically indistinguishable across the two treatments ($p = 0.26$, one-sided Fisher’s exact test).

These findings are confirmed in Tobit regressions in Table 2. We regress the number of punishment points a subject i receives in a period on the average contribution of the other group members and i 's deviation from this average, allowing for different coefficients for positive and negative deviations, and also controlling for period effects. Both with and without incentives, the coefficient on negative deviations from the others' average is negative and highly significant, meaning that the farther a subject's contribution is below the average, the more the subject is punished. However, as can be seen in the last column, this effect is significantly stronger without incentives, confirming the impression from the graph. In terms of magnitudes, the predicted marginal effects are that an additional one-point negative deviation from the average leads to a 0.315 point increase in punishment in the BT but only a 0.217 point increase in the IT.²¹ Thus, the marginal effect of negative deviations on punishment is about 31 percent lower in the IT than in the BT.

Positive deviations from the average, on the other hand, do not significantly affect received punishment in either treatment. A higher average contribution of other group members is predicted to significantly reduce punishment in the BT but not in the IT (however, the predicted marginal effect is small in the BT — a one-point increase in others' average contribution is predicted to reduce punishment received by 0.06 points). Finally, note that in the pooled regression, the incentive dummy is negative and significant, meaning that there is less punishment in the IT, controlling for the others' average and deviations from the average.²²

Another way to appreciate the quantitative difference between the two treatments in the intensity of punishment of free riders is to look at the predicted punishment

²¹The marginal effects refer to changes in the unconditional expected number of punishment points received, and are calculated from the "pooled" regression in the final column of the table, at the sample means for all values. The marginal effects predicted from columns (1) and (2) are very similar.

²²However, this coefficient may simply compensate for the absence of a negative coefficient on others' average in the IT.

received by a hypothetical subject. Assume a subject contributes 10 ECU while the three other group members contribute on average 15 ECU. The coefficients from the Tobit regression then predict that the subject receives 2.28 punishment points in the BT, but only 1.57 punishment points in the IT. If the hypothetical subject contributes nothing, he is predicted to receive 11.92 punishment points in the baseline and only 7.59 in the IT.

4.3 The Reaction of Free Riders to Punishment

RESULT 3: *For a given level of punishment, free riders increase their subsequent contributions by less in the IT than in the BT. This seems mostly due to the unwillingness of free riders in the IT to increase their contribution towards the average of their fellow group members.*

Again, we provide both graphical and regression support for the result. Figure 4 displays the changes in contributions from one period to the next for free riders (that is, subjects who contributed less in the previous period than the other group members did on average) who received different numbers of punishment points. Clearly, free riders increase their contributions by more if they are punished more heavily. However, the figure also shows that free riders tend to increase their contributions by less in the IT than in the BT. On average, free riders in the BT increase their contribution in the next period by 3.91 points, while in the IT, the increase is only 2.02 points, and this difference is statistically significant at $p < 0.01$ (Mann-Whitney test, $z = 2.767$). Looking only at those free riders who were punished, the average increases were 4.61 and 2.80 points ($z = 2.190, p < 0.03$).

In order to disentangle what drives the differences in the behavior of free riders between the two treatments, we look at regressions of the change in contribution on the

severity of punishment and other explanatory variables such as the average contribution of other group members in the previous period and the free rider's deviation from it.

The first column of Table 3 shows the coefficients of a regression of the change in contribution, $g_i^t - g_i^{t-1}$, on the cost of received punishment in the previous period, C_i^{t-1} , and the deviation of the subject's contribution in the previous period from the average of the other subjects' contributions, $g_i^{t-1} - \bar{g}_{-i}^{t-1}$ (which are all negative, given that we include only free riders in the regression). We interact the explanatory variables with an IT dummy, to detect differences between the two treatments. As in the previous section, we also control for period effects by including period dummies.

The results of the first column show that free riders increase their contributions significantly more the more heavily they were punished in the previous period. The coefficient of 0.156 means that for each punishment point a free rider receives, he or she on average increases his subsequent contribution by about 0.47 points, as each punishment point received costs him 3 ECU (unless he is punished so heavily that his whole first-stage payoff is lost). The coefficient on the interaction of the incentive dummy and the cost of punishment is very small and insignificant, meaning that the marginal effect of punishment on subsequent contribution changes is equally strong in both treatments. However, the coefficient on the incentive dummy itself is negative and mildly significant ($p = 0.056$), meaning that for a given level of punishment and deviation from the average, free riders in the IT tend to increase their contributions by less than free riders in the BT do. On the other hand, in neither treatment does it seem to matter how much below the others' average a free rider's contribution was in the previous period. This is somewhat surprising, since, for instance, Masclet et al. (2003) find a large and significant coefficient on this variable when running a similar regression.

The second column uses the other group members' average contribution, \bar{g}_{-i}^{t-1} , as

an explanatory variable, instead of i 's deviation from this average. The coefficient on the cost of punishment is slightly lower but even more highly significant, and there is still no significant difference across the two treatments. However, the coefficient on the incentive dummy has switched sign and is now positive. This is because the coefficient on others' average contribution is strongly significantly positive for the BT but very close to zero for the IT. Thus, the difference between the two treatments can be explained by the fact that in the BT, free riders increase their contribution more the more other group members contributed in the previous period, while the same is not true in the IT.²³

Column (3) confirms these findings by re-introducing the free rider's lagged negative deviation from the average as an explanatory variable, which does still not enter the regression significantly, and leaves the other coefficients largely unchanged as compared with the results shown in column (2).²⁴

²³To assess the difference between the two treatments in this regression, it is helpful to consider whether the predicted contribution change absent any punishment is higher in the BT than in the IT as a function of others' average contribution. It turns out that the predicted contribution change is significantly higher (at $p \leq 0.05$) for $\bar{g}_{-i}^{t-1} \geq 11.2$, which is the case for more than 70 percent of free riders. At the median value of \bar{g}_{-i}^{t-1} , which is 14.5, the regression predicts that a free rider in the IT changes his contribution upwards by 2.11 points less than a free rider in the IT.

²⁴Due to the negative (but insignificant) coefficient on the interaction of the incentive dummy and $g_i^{t-1} - \bar{g}_{-i}^{t-1}$, the exercise conducted in the previous footnote now leads to predicting somewhat smaller and less significant differences between the two treatments. At median values for free riders of the two explanatory variables other than cost of punishment (which is again assumed to equal zero), the regression coefficients from column (3) predict that a free rider in the IT changes his contribution upwards by 1.52 points less than a free rider in the IT, and the p-value of this difference is $p = 0.085$. If we drop the period dummies (which are all insignificant) and the interaction of incentive and cost from punishment, the predicted difference is 1.82 points and is significant at $p < 0.01$.

5 Discussion

5.1 Interpretation of Results

The main findings discussed in the previous section support the hypothesis that incentives and norm enforcement are *substitutes*, meaning that one or the other in isolation is successful in raising contributions, while adding incentives in a setting with a peer punishment mechanism does not lead to higher contributions. We find this to be due to two effects: (1) free riders receive significantly less punishment when incentives to contribute are present, and (2) they increase their subsequent contributions by less, whether or not they are punished.

Our preferred interpretation of the first effect is that the rewards (in the form of lottery tickets) received by the high contributors dampen their anger towards the free riders, resulting in less punishment. Thus, in terms of our discussion in Section 3.1, we find support for “Case 3”, namely, that lessening the inequality of outcomes by providing incentives reduces the anger of high contributors. The second effect is likewise consistent with the idea that the presence of contribution incentives mitigates the shame or guilt of free riders.

However, it may be debatable to what extent either (or both) of these explanations for the two effects is more compelling than explanations based on strategic reasoning. In particular, it is possible that the reduction in punishment is due to punishers’ anticipating that free riders will not adjust their contributions upwards very much. Similarly, free riders may react less strongly to punishment in the IT than in the BT not because they feel less shame or guilt, but because they anticipate that high contributors will not punish them very harshly if they keep contributing less than average. While our data do not allow us to rule out these “strategic” explanations for what we observe, we can look at what happens in and after the first period when the punishment mechanism

is available to the agents. If less punishment was inflicted on free riders in the IT because these free riders react less strongly to punishment than when no incentives are available, we might expect that in the first period punishment severity would be similar in both conditions.²⁵ Yet, in our data, a free rider in the BT receives on average 6.2 punishment points in the first period, while in the IT, the corresponding average is only 3.7 points (Mann-Whitney test: $z = 2.620, p < 0.01$). Unless high contributors in the IT somehow foresee that punishing free riders is “not worth it,” which we think is unlikely, this observation means that “dampened anger” seems to be a better explanation than strategic considerations for the less harsh punishment of free riders in the IT. This interpretation is also in line with previous research, mentioned in Section 3.1, which finds that strategic explanations have rather low explanatory power for punishment behavior in such experiments.

As for the reaction of free riders, it is harder to dismiss the possibility that the weaker reaction of free riders (whether or not they are punished) is due to strategic reasoning, particularly because, as just mentioned, punishment is less harsh in the IT from the beginning. Thus, even though the contribution increases of free riders from one period to the next are lower in the IT from the beginning (they increase their contribution by 2.9 points on average between the first and the second period of the punishment condition, while the corresponding number in the BT is 5.7 points (Mann-Whitney $z = 1.91, p < 0.06$)), we do not know whether this is a result of reduced guilt/shame or whether they anticipated that failing to increase their contributions would not result in harsh punishment.

Another alternative explanation for what we observe is that the presence of incentives makes contributing seem less a “social act” and more an individual choice motivated at least partially by private benefits; this could also explain why free riders

²⁵Then, over time the (potential) punishers would realize that the free riders do not react much to punishment, and would reduce their punishment accordingly.

are not induced to increase their contributions towards the mean in the IT, while in the BT they are. A free rider in the IT may think that his fellow group members contribute a lot only because they are after the lottery tickets, not because they genuinely care about the well-being of the group, and therefore may feel no compunction about contributing less. In addition, such a subject may feel that punishment he receives from high contributors is unjustified, and he may therefore refrain from increasing his subsequent contributions out of spite or principle. Again, we cannot rule out this explanation as an alternative to our story, which focuses more directly on the effect of incentives on emotions as motivators of behavior. However, we believe that this explanation fails to explain why high contributors punish free riders less harshly, unless one assumes that high contributors to some extent engage in “self-signaling” and infer their own motivation from their actions and the environment.²⁶ Furthermore, in the condition with no punishment mechanism, the lottery tickets perform well in increasing contributions, so there is no indication that they directly crowd out subjects’ intrinsic motivation to contribute.

5.2 A Word on Welfare

Even though the private incentives provided in the IT fail to lead to higher contributions when peer punishment is possible, this does not mean that they are not welfare-enhancing. This is because achieving a certain contribution level with less peer punishment is a good thing, as punishment is socially wasteful. Furthermore, the subjects in our IT are better off than the ones in the BT because we give them a lottery ticket for each point they contribute.²⁷ However, to assess the overall welfare effect of intro-

²⁶Then, for any given difference in contribution between two group members, the one who contributed more might feel relatively less strongly in the IT that he is being more prosocial than the other than in the BT.

²⁷The mean expected payoff (taking the value of a lottery ticket to be 0.2 ECU, its expected value) of a subject in the punishment condition of the IT was 38.0 ECU, as compared with 33.5 ECU in the

ducing incentives, one must take into account the cost of providing them, so the value of the lottery tickets should not enter the welfare calculation.²⁸ Therefore, we use expression (2) to compare welfare across treatments. Using this criterion, mean welfare is slightly, but not significantly, higher in the punishment condition of the IT than in the punishment condition of the BT.²⁹ This is mostly due to the first two periods, during which contributions are higher in the IT than in the BT and the still numerous free riders in the IT are punished harshly. Looking at periods 3 to 6 only, mean welfare is almost equal across the two treatments (36.4 ECU in the BT versus 36.0 ECU in the IT).³⁰

Generally, harsh punishment of free riders, which has a high social cost in the short run, can be expected to lead to high contributions in the long run mainly due to its effect as a threat, not because it is actually exercised (see, for example, Fehr and Gächter, 2000). Thus, over time, we would expect a decline in the differences in welfare costs due to punishment across treatments. Meanwhile, the private incentives must be provided in each period, and if their provision is socially costly (for instance because of deadweight losses arising in their financing), then welfare may be higher without them. Of course, we cannot claim from our experimental results that this is what would actually happen in reality, but we believe that this possibility, which arises as a result of the detrimental effect of incentives on the effectiveness of the norm enforcement mechanism, should at least be considered by a policymaker or manager in

BT (Mann-Whitney $z = 4.55, p < 0.001$)

²⁸In the experiment, we (the experimenters) finance the incentive. However, in real world situations, it would have to be financed through taxes (in case of incentives provided by the government) or directly by the party that is interested in raising contributions. This might lead to an additional welfare cost (because of deadweight losses from taxation, for instance).

²⁹The respective mean payoffs not including lottery tickets are 35.0 ECU in the IT and 33.5 ECU in the BT (Mann-Whitney $z = 1.42, p > 0.15$).

³⁰As mentioned earlier, mean contributions are slightly higher in the BT during these periods; also, total punishment during these periods is also almost the same in the two treatments. However, it is important to note that the result discussed in Section 4.2, namely, that free riders are punished more severely in the BT than in the IT, still holds for these periods — there is no difference in total punishment because there are fewer free riders in the BT.

deciding whether to introduce (additional) private incentives for prosocial behavior.

6 Conclusion

In our laboratory public good experiment, we find that private incentives for prosocial behavior, which substantially increase contributions in the condition without norm enforcement, fail to do so when norm enforcement is possible. This is due to the effects of the incentives on the severity with which free riders are punished, and on free riders' subsequent reaction. Our preferred interpretation of these findings is that being rewarded for their contributions reduces the anger of high contributors towards free riders, and that free riders may feel less shame or guilt for failing to contribute their "fair share."

Thus, we have identified another mechanism through which incentives can have unintended side-effects, so-called "hidden costs." While the existing literature has identified several ways in which monetary incentives could *directly* crowd out prosocial behavior, our finding can instead be seen as an *indirect* hidden cost, as it operates through reduced effectiveness of the norm enforcement mechanism. As such, it should be of concern for policymakers and managers who contemplate introducing private monetary incentives in settings where norm enforcement can be expected to play a significant role in generating high contributions to the public good.

Several questions related to our hypotheses and findings await further research. For instance, it would be interesting to know more about exactly what determines the strength of the negative emotions towards free riders, which in turn trigger punishment. In our interpretation, the presence of incentives weakens these negative emotions. This interpretation is consistent with inequality of outcomes as a driving force behind the negative emotions. It would be desirable to elicit these emotions more directly, either

through questionnaires (as done, for instance, by Hopfensitz and Reuben, 2009) or through physiological or neuroscientific measurement. Also, we believe that the framing of the incentives may be important for their effect on norm enforcement. We chose to make the incentive very salient, but it may be that results would be quite different if we had instead implemented a direct rebate, such that the cost of contributing decreases without an explicit reward for contributing. Likewise, it would be interesting to see what would happen if instead of rewards for contributing, fines for not contributing were introduced.

References

- ANDREONI, J. AND J. MILLER (2002): “Giving according to GARP: An experimental test of the consistency of preferences for altruism,” *Econometrica*, 70, 737–753.
- ARIELY, D., A. BRACHA, AND S. MEIER (2009): “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially,” *American Economic Review*, forthcoming.
- AUTEN, G., H. SIEG, AND C. T. CLOTFELDER (2002): “Charitable Giving, Income and Taxes: An Analysis of Panel Data,” *American Economic Review*, 92, 371–382.
- BENABOU, R. AND J. TIROLE (2006): “Incentives and Prosocial Behavior,” *American Economic Review*, 96, 1652–1678.
- BOSMAN, R. AND F. VAN WINDEN (2002): “Emotional Hazard in a Power-to-Take Experiment,” *Economic Journal*, 112, 147–169.
- BOWLES, S. (2008): “Policies Designed for Self-Interested Citizens May Undermine “The Moral Sentiments”: Evidence from Economic Experiments,” *Science*, 320, 1605–1609.
- BOWLES, S. AND H. GINTIS (2005): “Prosocial Emotions,” in *The Economy as a Complex Evolving System III: Essays in Honor of Kenneth Arrow*, ed. by L. Blume and S. Durlauf, Oxford University Press, Oxford.
- BRANDTS, J. AND G. CHARNES (2003): “Truth or consequences: An experiment,” *Management Science*, 49, 116–130.
- CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2006): “Robust Inference with Multi-way Clustering,” Technical Working Paper 327, National Bureau of Economic Research.
- CARPENTER, J., S. BOWLES, H. GINTIS, AND S.-H. HWANG (2009): “Strong Reciprocity and Team Production: Theory and Evidence,” *Journal of Economic Behavior and Organization*, forthcoming.
- DAWES, C. T., J. H. FOWLER, T. JOHNSON, R. MCELREATH, AND O. SMIRNOV (2007): “Egalitarian Motives in Humans,” *Nature*, 446, 794–796.
- DE QUERVAIN, D. J.-F., U. FISCHBACHER, V. TREYER, M. SCHELLHAMMER, U. SCHNYDER, A. BUCK, AND E. FEHR (2004): “The Neural Basis of Altruistic Punishment,” *Science*, 305, 1254–1258.
- DECI, E. L. (1975): *Intrinsic Motivation*, New York: Plenum Press.

- DECI, E. L., R. KOESTNER, AND R. M. RYAN (1999): “A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation,” *Psychological Bulletin*, 125, 627–668.
- EGAS, M. AND A. RIEDL (2008): “The economics of altruistic punishment and the maintenance of cooperation,” *Proceedings of the Royal Society B: Biological Sciences*, 275, 871–878.
- ELLINGSEN, T. AND M. JOHANNESSON (2008): “Pride and Prejudice: The Human Side of Incentive Theory,” *American Economic Review*, 98, 990–1008.
- FALK, A., E. FEHR, AND U. FISCHBACHER (2005): “Driving Forces Behind Informal Sanctions,” *Econometrica*, 73, 2017–2030.
- FALK, A. AND M. KOSFELD (2006): “The Hidden Cost of Control,” *American Economic Review*, 96, 1611–30.
- FEHR, E. AND A. FALK (2002): “Psychological Foundations of Incentives,” *European Economic Review*, 46, 287–324.
- FEHR, E. AND S. GÄCHTER (2000): “Cooperation and Punishment in Public Goods Experiments,” *American Economic Review*, 90, 980–994.
- (2002): “Altruistic punishment in humans,” *Nature*, 415, 137–140.
- FEHR, E. AND J. A. LIST (2004): “The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs,” *Journal of the European Economic Association*, 2, 743–71.
- FISCHBACHER, U. (2007): “z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental Economics*, 10, 171–178.
- FOWLER, J. H., T. JOHNSON, AND O. SMIRNOV (2004): “Egalitarian motive and altruistic punishment,” *Nature*, 433, E1–E2.
- FREY, B. S. (1997): *Not Just for the Money. An Economic Theory of Personal Motivation*, Cheltenham, UK and Brookfield, USA: Edward Elgar.
- FREY, B. S. AND R. JEGEN (2001): “Motivation Crowding Theory: A Survey of Empirical Evidence,” *Journal of Economic Surveys*, 5, 589–611.
- FREY, B. S. AND F. OBERHOLZER-GEE (1997): “The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out,” *American Economic Review*, 87, 746–55.
- GNEEZY, U. AND A. RUSTICHINI (2000a): “A Fine Is a Price,” *Journal of Legal Studies*, 29, 1–18.

- (2000b): “Pay Enough or Don’t Pay at All,” *Quarterly Journal of Economics*, 115, 791–810.
- HERRMANN, B., C. THÖNI, AND S. GÄCHTER (2008): “Antisocial Punishment Across Societies,” *Science*, 319, 1362–1367.
- HEYMAN, J. AND D. ARIELY (2004): “Effort for Payment: A Tale of Two Markets,” *Psychological Science*, 15, 787–93.
- HOPFENSITZ, A. AND E. REUBEN (2009): “The Importance of Emotions for the Effectiveness of Social Punishment,” *Economic Journal*, forthcoming.
- LAURY, S. K. (2005): “Pay One or Pay All: Random Selection of One Choice for Payment,” *Working Paper*.
- LEDYARD, J. O. (1995): “Public Goods: A Survey of Experimental Research,” in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton, NJ: Princeton University Press, 111–194.
- LEPPER, M. R. AND D. GREENE, eds. (1978): *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*, Hillsdale, NY: Erlbaum.
- LEVINE, D. K. (1998): “Modeling Altruism and Spitefulness in Experiments,” *Review of Economic Dynamics*, 1, 593–622.
- MAS, A. AND E. MORETTI (2009): “Peers at Work,” *American Economic Review*, forthcoming.
- MASCLET, D., C. NOUSSAIR, S. TUCKER, AND M.-C. VILLEVAL (2003): “Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism,” *American Economic Review*, 93, 366–380.
- MASCLET, D. AND M.-C. VILLEVAL (2008): “Punishment and Inequality: A Public Good Experiment,” *Social Choice and Welfare*, 31, 475–502.
- MELLSTRÖM, C. AND M. JOHANNESSON (2008): “Crowding Out in Blood Donation: Was Titmuss Right?” *Journal of the European Economic Association*, 6, 845–863.
- OSTROM, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
- REUBEN, E. AND A. RIEDL (2009): “Public Goods Provision and Sanctioning in Privileged Groups,” *Journal of Conflict Resolution*, 53, 72–93.
- VYRASTEKOVA, J., Y. FUNAKI, AND A. TAKEUCHI (2008): “Strategic vs. Non-Strategic Motivations of Sanctioning,” *Working Paper*.

Table 1: Treatments

	Baseline Treatment (Without Private Incentives)	Incentive Treatment (With Private Incentive)
Without Punishment (six periods)	15 groups of size 4	19 groups of size 4
With Punishment (six periods)	15 groups of size 4	19 groups of size 4

Table 2: Determinants of Punishment Points Received

	(1) Without Incentive	(2) With Incentive	(3) Pooled
Incentive			-3.038** (1.514)
Neg. deviation from others' avg.	-1.053*** (.076)	-.718*** (.112)	-1.032*** (.072)
Incentive x Neg. dev.			.32*** (.12)
Pos. deviation from others' avg.	-.137 (.089)	-.014 (.119)	-.174 (.111)
Incentive x Pos. dev.			.165 (.165)
Others' average contribution	-.201*** (.078)	.037 (.063)	-.197** (.087)
Incentive x Others' avg.			.233** (.108)
Constant	-.45 (1.552)	-2.462* (1.26)	.21 (1.341)
Period dummies	Yes	Yes	Yes
# of observations	360	456	816

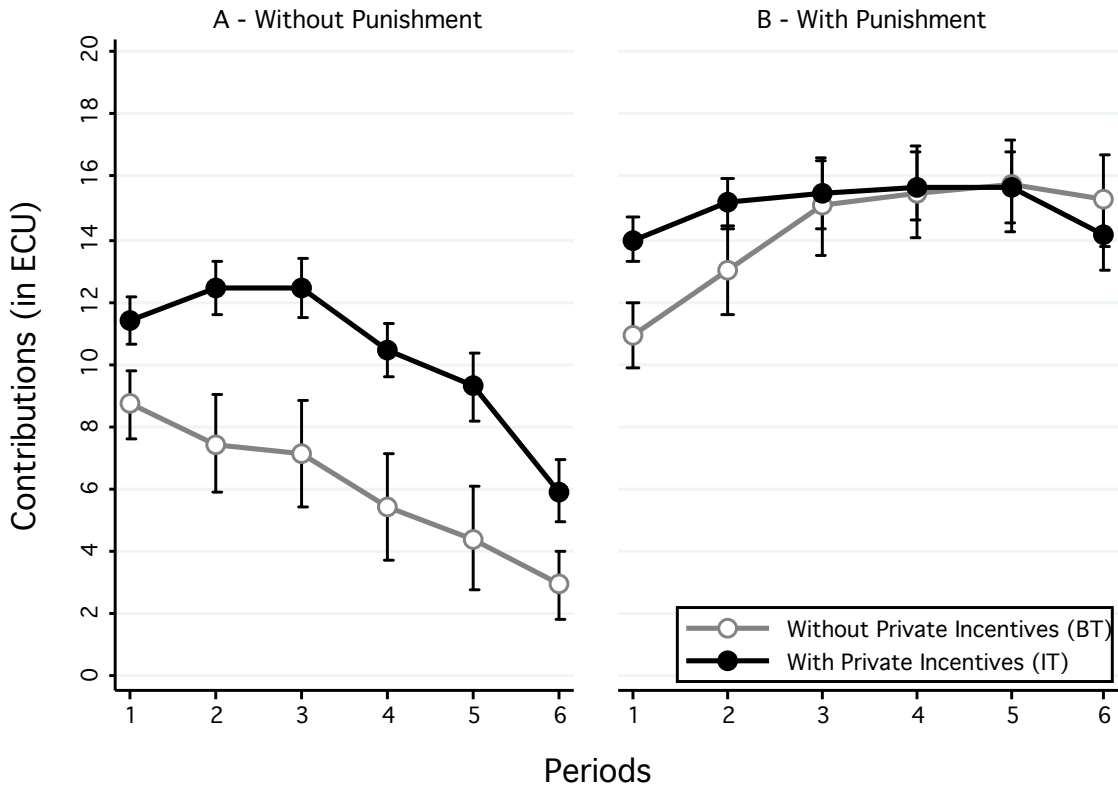
Note: Dependent variable: Punishment points received by a subject. Tobit regressions. "Incentive" is a dummy variable that equals one for observations from the Incentive Treatment. "Neg. deviation from others' avg." = $\min(0, g_i - \bar{g}_{-i})$; "Pos. deviation from others' avg." = $\max(0, g_i - \bar{g}_{-i})$. Standard errors in parentheses clustered at the group level.

Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Determinants of Free Riders' Contribution Changes

	(1)	(2)	(3)
Incentive	-1.489* (.78)	2.201 (1.529)	1.448 (1.743)
Cost of punishm. received last pd	.156** (.07)	.134*** (.049)	.17*** (.062)
Incentive x Cost of punishm.	-.013 (.099)	.064 (.07)	-.024 (.098)
Neg. deviation from others' avg. last pd	-.051 (.22)		.124 (.25)
Incentive x Neg. dev. last pd	-.132 (.266)		-.305 (.29)
Others' average contribution last pd		.263*** (.071)	.294*** (.101)
Incentive x Others' avg. last pd		-.297*** (.109)	-.31** (.135)
Constant	.99 (.618)	-1.463 (.99)	-1.642 (1.028)
Period dummies	Yes	Yes	Yes
# of observations	218	218	218

Note: Dependent variable: Change in contribution. Linear (OLS) regressions. Standard errors in parentheses clustered at the group and the individual level, following Cameron et al. (2006) and using their 'cgmreg' routine in Stata.
Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



Note: Bars show standard errors of the mean.

Figure 1: Effect of Incentives on Mean Contributions to Public Good

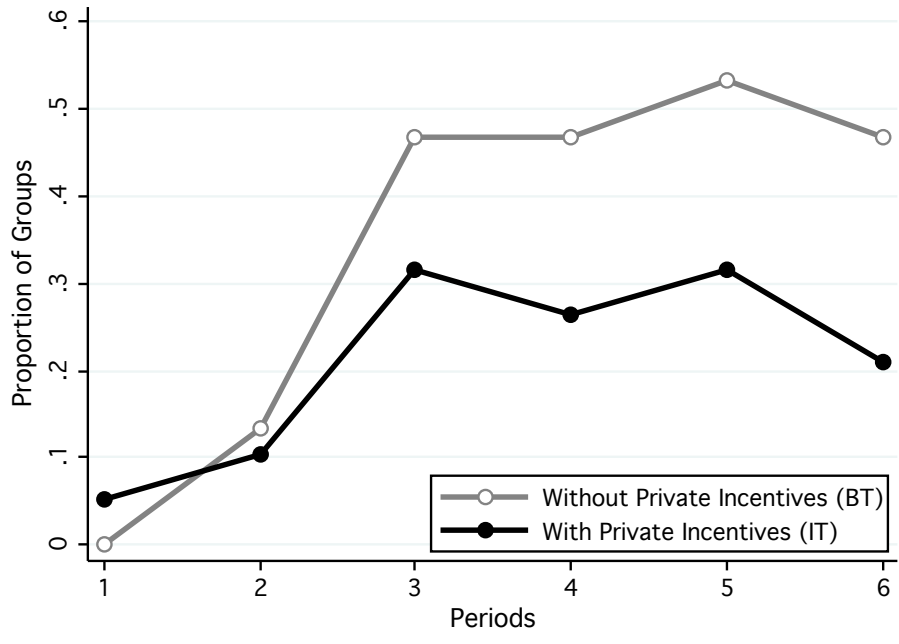
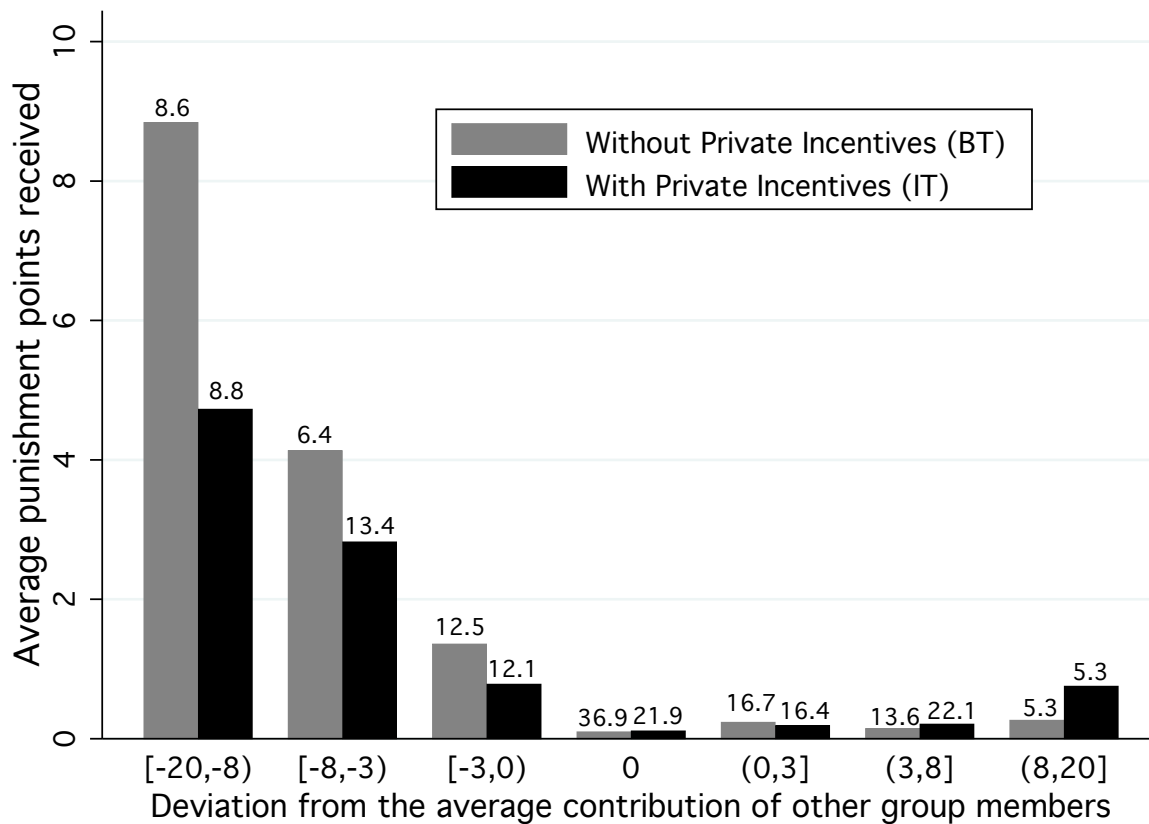
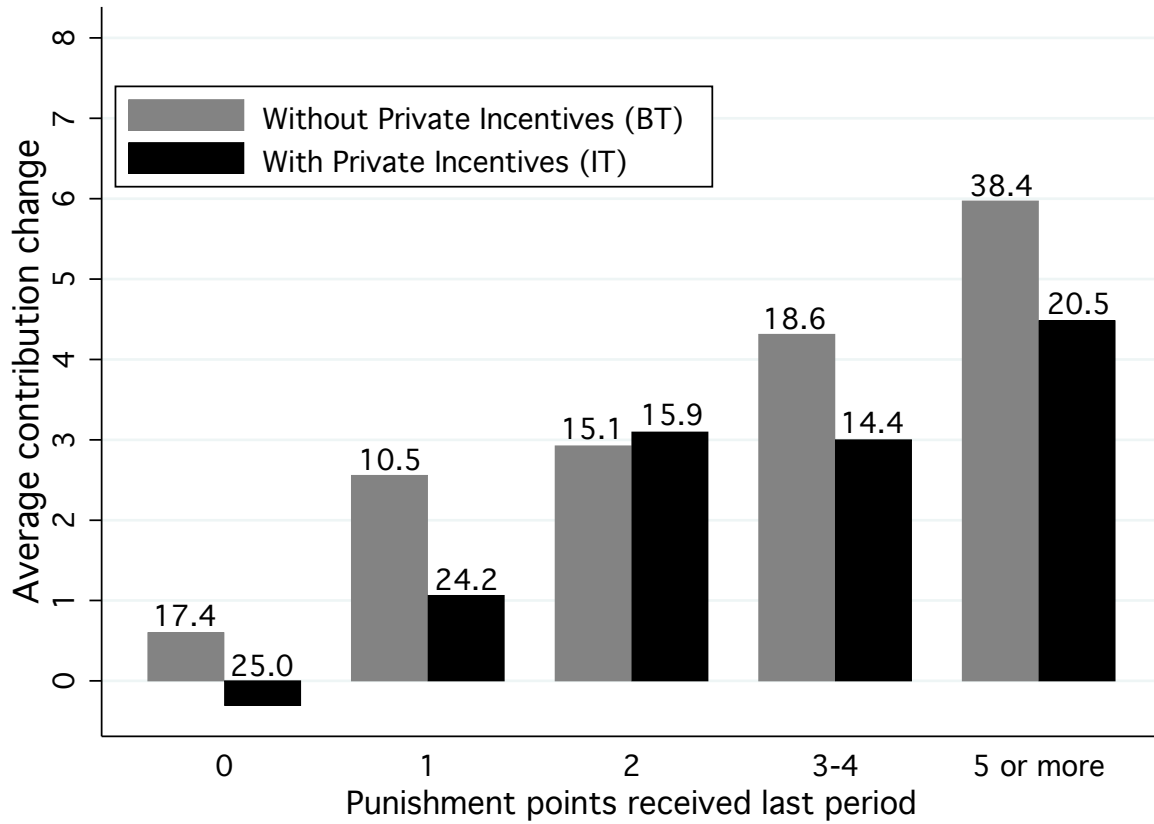


Figure 2: Proportion of Groups Reaching Socially Efficient Contribution Level in the Punishment Condition



Note: Numbers above bars indicate the relative frequency of observations in a category.

Figure 3: Received Punishment Points for Deviations from Others' Average Contribution



Note: Numbers above bars indicate the relative frequency of observations in a category.

Figure 4: Contribution Change of Free Riders in Punishment Condition

Appendix: Instructions

[The following instructions are for the two parts of an Incentive Treatment Session. Instructions for the Baseline Treatment were identical, except that all references to “lottery points” were removed.]

Instructions for Participants

You are now taking part in an economic experiment. If you read and follow these instructions carefully, you can, depending on your decisions and the decisions made by other participants, earn a considerable amount of money. It is therefore important that you take the time to understand the instructions.

Before we begin, we ask you to respect the following guidelines:

- Please, no talking. If you have any questions during the study, please raise your hand. An assistant will come to your place and answer your question privately.
- Every participant’s task is individual and should be completed in private. No one except you should be looking at your screen at any time.
- Please turn off your cell phone.
- Do not use the computer for any purpose besides the study. Exiting the program or running other applications may compromise the study and the data it provides. When you are finished, please wait quietly until given further instructions.

If you do not comply with these rules, we will be forced to exclude you from the study and you will not receive any money. Thank you for your cooperation!

There will be two separate experiments. The two experiments are independent of each other and the details of them will be explained one after the other.

During the experiments we will not speak of dollars, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECUs. At the end of the experiments, the amount of ECUs you have earned will be converted to dollars at the rate of

$$1 \text{ ECU} = \$ 0.25.$$

After the experiments your earnings will be paid to you **in cash, together with the \$10 show-up fee.**

Instructions for Experiment # 1

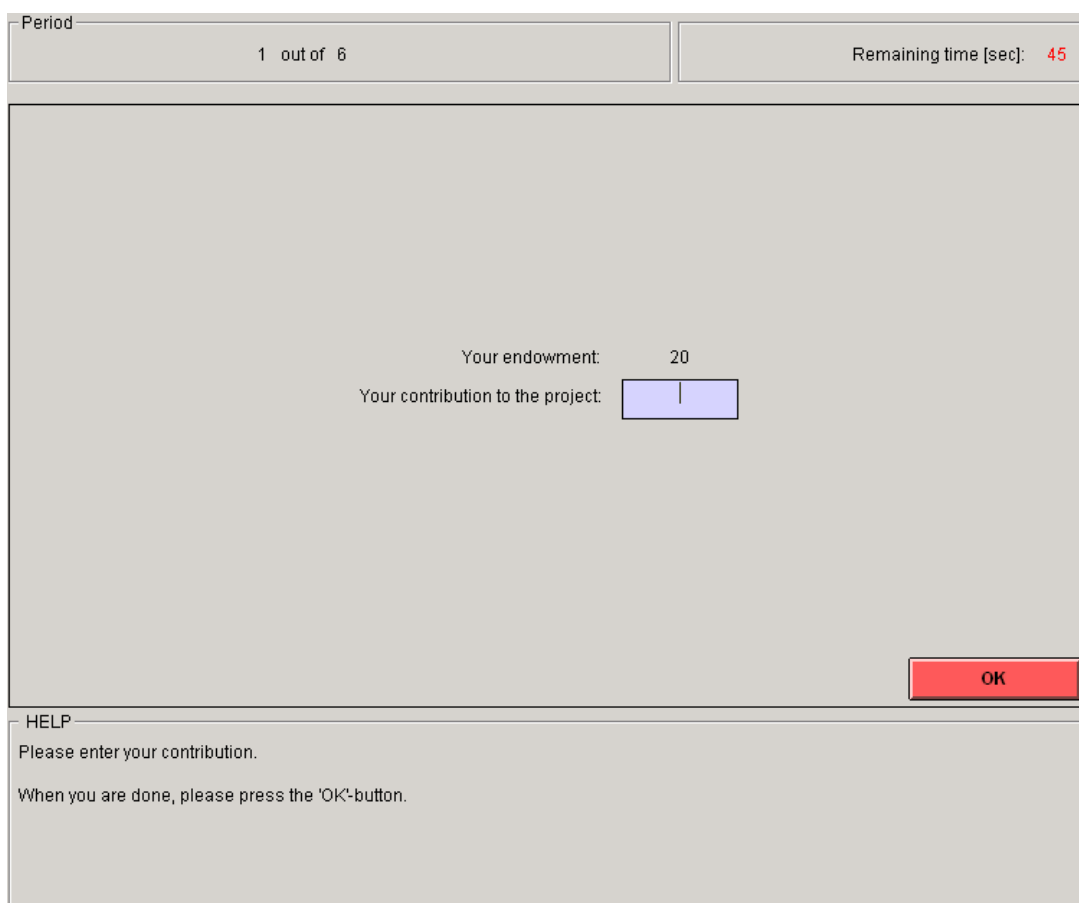
This experiment lasts 6 periods. **At the end of the experiments, one of these 6 periods will be randomly selected for payment.**

At the beginning of the experiment, the participants will be randomly divided into groups of four. You will therefore always be in a group with three other participants. The groups will remain the same throughout this experiment. This means that you will always interact with the same three participants in all 6 periods.

We will now describe what happens in each period.

At the beginning of every period, each participant will receive **20 ECUs**. In the following, we will refer to this amount as the “**endowment.**” Your task is to decide how to use your endowment. In particular, **you must decide how many of the 20 ECUs you will contribute to a project** (from 0 to 20) **and how many of them you will keep for yourself.** The consequences of your decision are explained below.

At the beginning of each period, you will see the following decision screen where you can enter your decision:



Period 1 out of 6 Remaining time [sec]: 45

Your endowment: 20
Your contribution to the project: 1

OK

HELP
Please enter your contribution.
When you are done, please press the 'OK'-button.

In the top left corner, you can see which period you are in. In the top right corner, the remaining time for you to make your decision is displayed (in seconds). You need to make your decision before the counter reaches zero seconds.

To register your decision, you enter the number of ECUs that you intend to contribute to the project (a whole number between 0 and 20) in the field provided and then press the “OK” button.

Once all the players have chosen their contributions to the project, the payoff screen (shown below) will inform you of the group's total contribution (including your contribution), your income from the project, and your payoff in this stage.

Period	1 out of 6	Remaining time [sec]: 36
Your contribution to the project: ...		
Sum of all contributions to the project: ...		
Income from the points that you kept: ...		
+ Income from the project: ...		
= Your income in this period: ...		
Lottery points earned in this period: ...		
<input type="button" value="Continue"/>		
HELP		
This screen shows the outcome of this period.		
Once time is up or once everybody has pressed the 'Continue'-button, the experiment will continue.		

As you can see, your income at the end of the period consists of three parts:

- 1) The ECUs that you kept for yourself (“Income from the points that you kept”); and
- 2) The income from the project (“Income from the Project”), which is calculated as follows:

$\text{Your income from the project} = 0.4 * \text{Sum of all contributions to the project}$
--

- 3) A number of “lottery points,” equal to the number of ECUs that you contributed to the project. At the end of the experiment, we will conduct a lottery, in which each participant is eligible to win a prize of 20 ECU. Your probability of winning will be determined by the number of lottery points

that you earned in the period that is chosen for payment. Your probability of winning depends only on your own decisions, not on the decisions of others. If you earned x lottery points in the selected period, your probability of winning the 20 ECU would be $x\%$. For instance, if your contribution to the project is 0, your chance of winning 20 ECU in the lottery is 0%. If your contribution to the project is 15, your chance of winning 20 ECU in the lottery is 15%, etc.

Your payoff for a period is therefore calculated using the following formula:

$\text{Income in this period (in ECU)} = (20 - \text{Your contribution to the project}) \quad (1)$
$+ 0.4 * (\text{Sum of all contributions to the project}) \quad (2)$
In addition, you earn x Lottery points (where $x = \text{Your contribution to the project}$).

The income of each group member from the project (Part (2) in the box above) is calculated in the same way. This means that each group member receives the same income from the project.

Suppose the sum of the contributions of all group members is 60 ECUs. In this case, each member of the group receives an income from the project of: $0.4 * 60 = 24$ ECUs. If the total contribution to the project is 9 ECUs, then each member of the group receives an income of: $0.4 * 9 = 3.6$ ECUs from the project, regardless of how much the member individually contributed to the project. (There is a table at the end of these instructions to help you with these calculations.)

You always have the option of keeping the ECUs for yourself or contributing them to the project. Each ECU that you keep raises your period income by 1 ECU. If you contributed this ECU to the project instead, the total contribution to the project would rise by 1 ECU, and the probability of your winning the 20 ECU in the lottery would increase by 1 percentage point. Your income from the project would thus rise by $0.4 * 1 = 0.4$ ECU. However, the income of each of the other group members would also rise by 0.4 ECU, so the total income of the group from the project would increase by 1.6 ECUs. Thus, your contribution to the project also raises the income of the other group members. On the other hand, you also benefit from each ECU contributed by the other group members to the project: for each ECU contributed by any member you earn 0.4 ECU.

After the payoff screen, the individual contributions of each group member will be displayed on the members' contribution screen:

Period		1 out of 6			Remaining time [sec]: 107	
Endowments:	20	20	20	20		
Contributions to the project:		
Contributions as % of the endowment:		
Lottery points earned:		

HELP
This screen shows you the decisions of the other members of your group in this period.
Once you are done, please press the 'Continue'-button.

Your contribution to the project will always be shown in the first column in a blue font. The other three members' contributions in this period will be displayed in random order in the following three columns. This random order will change in each period.

In addition to the absolute amount of the contributions, the members' contribution screen also shows contributions as a percentage of the endowment, and the number of lottery points earned by each participant.

After this screen, the next period starts, and the process repeats until the end of period 6.

Please raise your hand if you have any questions.

Control Questionnaire

Before we start, please answer the following questions and write down all the steps of your calculation. This is to make sure that all participants understand the details of how the study works.

THERE IS A TABLE ON THE LAST PAGE OF THESE INSTRUCTIONS TO HELP YOU WITH THE CALCULATIONS!

1. Each group member has an endowment of 20 ECUs. Nobody (including you) contributes any ECUs to the project. What is:

- a. Your income at the end of the period (without lottery points)?
- b. The number of lottery points you earned?
- c. The income of the other group members (without lottery points)?
- d. The number of lottery points each other group member earned?

2. Each group member has an endowment of 20 ECUs. You contribute 20 ECUs to the project. All other group members contribute 20 ECUs each to the project. What is:

- a. Your income at the end of the period (without lottery points)?
- b. The number of lottery points you earned?
- c. The income of the other group members (without lottery points)?
- d. The number of lottery points each other group member earned?

3. Each group member has an endowment of 20 ECUs. The other three group members contribute together a total of 30 ECUs to the project. What is:

- a. Your income at the end of the period if you contribute 0 ECUs to the project (without lottery points)?
- b. Your income at the end of the period if you contribute 15 ECUs to the project (without lottery points)?

4. Each group member has an endowment of 20 ECUs. You contribute 8 ECUs to the project.

What is:

- a. Your income if the other group members together contribute a further total of 7 ECUs to the project?
- b. Your income if the other group members together contribute a further total of 22 ECUs to the project?
- c. How many lottery points did you earn in both cases?

Please raise your hand once you are done, so an assistant can check your answers.

The following table will help you with the calculations to determine your income from the project. Two examples to illustrate how the table is read: if the sum of all contributions (yours included) equals 17, then your income (and everybody else's) from the project equals 6.8. If the sum of all contributions (yours included) equals 59, then your income (and everybody else's) from the project equals 23.6.

Number	* 0.4
1	0.4
2	0.8
3	1.2
4	1.6
5	2
6	2.4
7	2.8
8	3.2
9	3.6
10	4
11	4.4
12	4.8
13	5.2
14	5.6
15	6
16	6.4
17	6.8
18	7.2
19	7.6
20	8
21	8.4
22	8.8
23	9.2
24	9.6
25	10
26	10.4
27	10.8
28	11.2
29	11.6
30	12
31	12.4
32	12.8
33	13.2
34	13.6
35	14
36	14.4
37	14.8
38	15.2
39	15.6
40	16

Number	* 0.4
41	16.4
42	16.8
43	17.2
44	17.6
45	18
46	18.4
47	18.8
48	19.2
49	19.6
50	20
51	20.4
52	20.8
53	21.2
54	21.6
55	22
56	22.4
57	22.8
58	23.2
59	23.6
60	24
61	24.4
62	24.8
63	25.2
64	25.6
65	26
66	26.4
67	26.8
68	27.2
69	27.6
70	28
71	28.4
72	28.8
73	29.2
74	29.6
75	30
76	30.4
77	30.8
78	31.2
79	31.6
80	32

Instructions for Experiment # 2

The experiment will now be repeated for another set of 6 periods with a number of changes, which will be explained below.

As in the first set of 6 periods, one period will be randomly chosen to determine your payoff. **After these 6 periods, the entire experiment will be over and you will get your final payoff, which will be calculated as:**

Payoff from one randomly chosen period from the first set of 6 periods
+Payoff from one randomly chosen period from the second set of 6 periods
+Possible prize from the two lotteries (one for the first experiment, one for this experiment)
= Total earned payment (converted into dollars at 1 ECU = \$ 0.25)
+ \$10 show-up fee.

In this experiment, you will be matched **with three new participants**. None of the members of your group from the first experiment will be in your group now. **You will interact with the same participants in all 6 periods of this experiment.**

The experiment now (in this second set of 6 periods) consists of **two stages** in each period.

The first stage is identical to the experiment you just participated in. Thus, in the first stage of each period you must decide how much of your 20 ECU endowment to contribute to a project (and how much to keep for yourself). The payoff in the first stage, which is a provisional payoff, will be calculated exactly as in the previous experiment. For each ECU you keep, you will earn one ECU. For each ECU you contribute to the project, you earn a lottery point, and you and all the other group members earn 0.4 ECU. Each point contributed by another group member to the project also increases your payoff by 0.4 ECU.

What changes in the new experiment?

First of all, you will be given a lump sum of 10 ECU in each period. These 10 ECU cannot be invested in the project. They are simply added to your payoff at the end of each period.

More importantly, **in this experiment, there will be a second stage which will follow after the payoff screen of stage 1.** This second stage will be described now.

The second stage:

In the second stage of each period you will learn how much each group member contributed individually to the project in the first stage. Then, you can **reduce or leave equal** the income of **each** member of your group by **assigning deduction points**. The other group members can likewise reduce your income if they wish. This is why the payoff of the first stage was referred to as a “provisional payoff.”

You will be asked to make a decision using the following decision screen:

Period	1 out of 6	Remaining time [sec]: 110		
Endowments:	20	20	20	20
Contributions to the project:
Contributions as % of the endowment:
Lottery points earned:
Your deduction point assignment in stage 2:	-	<input type="text"/>	<input type="text"/>	<input type="text"/>
No point assignment: 0 Deduction points: whole number between 1 and 10		Cost Calculation		
The costs of your deduction point assignment are: -----				
OK				
HELP Please enter your decision. Then, press the 'Cost Calculation' button (you can still change your point assignments after that). Once you are done, please press the 'OK'-button.				

On this screen you will be informed about each group member's contribution. Your contribution to the project will always be shown in the first column in a blue font. The other three members' contributions in this period will be displayed in random order in the remaining three columns. Again, this random order changes in each period.

You must decide whether to assign any **deduction points** to each group member (except yourself), and, if so, how many to assign. You must enter a number for each group member in the appropriate field. **You can distribute deduction points between 0 and 10 in increments of whole points.**

The consequences of your point assignments

Assigning deduction points affects your payoff as well as the payoff of the group member to whom you assign the deduction points:

- Assigning points is costly to you. The costs will be computed according to the following formula: **Cost to you of assigning deduction points (in ECU) = Total number of deduction points you assign.**

Thus, each point that you assign will cost you 1 ECU. For example, if you assign 2 points to one member, it will cost you 2 ECU; if you in addition assign 9 points to another member, it will cost you 9 ECU; if you assign 0 points to the last member of the group, it will cost you nothing. In total, you would have assigned 11 points and your costs would be 11 ECU (2+9+0).

The total costs to you of your point assignments can be calculated on the computer by pressing the button "Cost Calculation." Before you press the OK-button, you can always change your decision and calculate the total costs to you again.

- **Each deduction point assigned to another group member will reduce the income of that member by 3 ECU.** If you choose 0 points for a particular group member, you do not change his or her income. However, if you assign **one deduction point** (that is, you enter 1), this will reduce the income of the person you assign it to by **3 ECU**. If you assign **two points** (that is, you enter 2), this will reduce the income of the person you assign it to by **6 ECU**, etc.

Whether and by how much the income of a group member is reduced in total depends on the total number of deduction points this group member is assigned. For example, if somebody receives in total (from all other members) 3 deduction points, the income of this person will be reduced by 9 ECU. If somebody receives 4 deduction points, his or her income will be reduced by 12 ECU.

The lottery points that you earned in the first stage are unaffected by the second stage.

Your total income from the two stages is then calculated as follows:

<p>Total income (in ECUs) at the end of the period =</p> <p>= Lump-sum payment of 10 ECU</p> <p>+ Income from the 1st stage</p> <p>- 3 * Deduction points you receive</p> <p>- Total cost of assigning deduction points.</p> <p>-----</p> <p>Should Income from the 1st stage – 3*Deduction points you receive</p> <p>be negative, your total income at the end of the period will simply equal</p> <p>Lump-sum payment of 10 ECU - Total cost of assigning deduction points.</p>
--

Two examples illustrate this:

- Suppose that your income from the first stage is 20, that you were assigned 3 deduction points by the other members of your group, and that you assigned 2 deduction points to others yourself. Your income in this period is then

$$\begin{aligned} & 10 \text{ (lump-sum payment)} + 20 \text{ (income from 1st stage)} \\ & - (3*3) \text{ (deduction points received)} - 2 \text{ (cost of assigned deduction points)} \\ & = 19 \end{aligned}$$

- Suppose that your income from the first stage is 20, that you were assigned 9 deduction points by the other members of your group, and that you assigned 2 deduction points to

others yourself. Because 20 (your income from 1st stage) $- (3 \cdot 9)$ (deduction points received) is less than 0 , your income in this period is then

$$10 \text{ (lump-sum payment)} - 2 \text{ (cost of assigned points)} = 8$$

Please note that your income in ECUs at the end of a period can be negative, but that you can prevent this with certainty through your own decisions.

Once all group members have made their decisions, you are informed of the outcome of the second stage and of your period income on the following income screen:

Period	
1 out of 6	Remaining time [sec]: 35
Your income in stage 1: ...	
The costs of your deduction point assignments in stage 2: ...	
Deduction points received: ...	
Income reduction due to deduction points received: ...	
Lump-sum payment 10	
Your total income in this period is therefore: ...	
Lottery points earned: ...	
<input type="button" value="Continue"/>	
HELP	
This screen shows the outcome of the second stage.	
Once time is up or once everybody has pressed the 'Continue'-button, the experiment will continue.	

Then, the next period begins (with stage 1), and the process repeats until the end of period 6.

Please raise your hand if you have any questions.

Control Questionnaire

Please answer the following questions and write down all the steps of your calculation. This is to make sure that all participants understand the details of the study.

1. In the second stage, you assign the following deduction points to the three other members of your group: 9, 5, 0. How much is the total cost to you of assigning those points?

2. How much is the total cost to you if, in the second stage, you assign 0 deduction points to all group members?

3. By how much will your income be reduced if the other members of your group together assign a total of 0 deduction points to you in the second stage?

4. By how much will your income be reduced if the other members of your group together assign a total of 4 deduction points to you in the second stage?

5. How much will your **total income after the period** be if your income in the first stage was 34, the other group members assign a total of 10 deduction points to you in the second stage, and...

a) you assign a total of 0 deduction points to other group members yourself?

b) you assign a total of 3 deduction points to other group members yourself?

(Don't forget the lump sum payment of 10 in your calculations! – It may be a good idea to refer to the box on page 4 of these instructions for the calculation.)

(continued on next page)

6. How much will **your total income after the period** be if **your income in the first stage was 20**, the other group members assign a total of 10 deduction points to you in the second stage, and...

a) you assign a total of 0 deduction points to other group members yourself?

b) you assign a total of 3 deduction points to other group members yourself?

Please raise your hand once you are done, so an assistant can check your answers.