

JENA ECONOMIC RESEARCH PAPERS



2010 – 013

Committing to Incentives: Should the Decision to Sanction be Revealed or Hidden?

by

Charlotte Klempt
Kerstin Pull

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Jena Economic Research Papers 2010 - 013

Committing to Incentives: Should the Decision to Sanction be Revealed or Hidden?

Charlotte Klempt * Kerstin Pull ‡

March 2, 2010

Abstract

Sanctions are widely used to promote compliance in principal-agent-relationships. While there is ample evidence confirming the predicted positive incentive effect of sanctions, it has also been shown that imposing sanctions may in fact reduce compliance by crowding-out intrinsic motivation. We add to the literature on the hidden costs of control by showing that these costs are restricted to situations where principals ex ante *reveal* their decision to sanction low compliance. If this decision is not revealed and agents do not know whether they will be sanctioned or not in case of low compliance, we do not find evidence of crowding-out - not even in those cases where agents firmly believe that they will be sanctioned in case of low performance.

PsycINFO: 2360; 3000

JEL classification: C72; C91; D03

Keywords: Intrinsic Motivation; Monetary Incentives; Job Performance

1 Introduction

Standard economic models of agency assume that imposing sanctions or fines in case of low performance increases agent incentives and leads agents to choose higher effort levels. While there is evidence for the expected incentive effect (see e.g. Fehr and Gächter, 2000, 2002; Andreoni et al., 2003, Henrich et al., 2006), it has also been shown that sanctions may have detrimental effects and may crowd out intrinsic motivation (e.g. Fehr and Rockenbach, 2003; Gneezy and Rustichini, 2000a; Houser et al., 2008; Fehr and Falk, 2002). Much in line with the literature on sanctions, also for extrinsic rewards (e.g. piece rates, bonus payments)

*Max Planck Institute of Economics, Kahlaische Straße 10, 07745 Jena, Germany

†Eberhard Karls Universität Tübingen, Faculty of Economics and Business Administration, Nauklerstr. 47, 72074 Tübingen, Germany

‡Corresponding author. kerstin.pull@uni-tuebingen.de, phone +49 - 7071 - 29 - 78186, Fax +49 - 7071 - 29 - 5077

both, positive incentive effects and negative crowding out effects were repeatedly identified (see e.g. Lazear, 2000 for the former, and Gneezy and Rustichini, 2000b or Gneezy 2002 for the latter). The same is true for acts of control such as principals' restricting agents' choice set of activity levels (Falk and Kosfeld, 2006). Among the most prominent explanations for crowding out caused by extrinsic incentives (carrots as well as sticks) is the presumption that agents may interpret these as a signal of distrust (e.g. Dickinson and Villeval, 2004; Frey 1993) or as a signal of a low underlying work norm (Sliwka, 2007) leading agents to reduce their level of voluntary cooperation.

In both theoretical and experimental work, the signalling effects of extrinsic incentives appear to be a crucial factor that causes crowding out. To the best of our knowledge, however, it has not been analyzed as yet what happens if this signal is missing i.e. if agents *do not know* whether they will be sanctioned or not in case of low performance. The basic motivation of our analysis is to identify the impact of the signalling effects on crowding out. We ask: Does it make a difference if the sanction policy is revealed or hidden?

In the literature, there is ample evidence on the positive incentive effects of *ex post* (i.e. not previously announced) punishment (not least the ultimatum game exemplifies these positive effects of ex post sanctions, see Camerer, 2003 for a review). While this indeed suggests that it is not the sanction as such that decreases performance in principal-agent-relationships, but rather its pre-announcement, there is no study as yet directly comparing the effects of ex post punishment with those of an ex-ante commitment to sanction within one experimental design. Accordingly, Fehr and Falk (2002) who acknowledge the potential differential effects of ex-post as opposed to ex-ante punishment, but at the same time state that these are not well studied as yet, "because this would require the conduct of an experiment with identical ex post and ex ante punishment opportunities".

In our paper, we fill this gap in the literature by explicitly comparing two sanction schemes where a principal's sanction policy is either revealed or hidden within one experimental design: When a principal's sanction decision is *revealed*, she commits herself ex-ante to either sanction low performance or not. *Hidden* sanctions, on the other hand, are based on ex-post retrospective punishment. Here, the principal decides whether she wants to sanction low performance or not without this decision being conveyed to the agent. As a control treatment we also regard a situation where there is *no* sanction option at all.

With the help of our experimental design, we are able to show that it is indeed the *ex-ante commitment* to sanction low performance that has detrimental crowding out effects on agent performance. We do not find these detrimental effects in case of hidden sanctions, i.e. when the sanctioning decision of the principal is not conveyed to the agent. Hidden sanctions lead to *higher* performance levels as compared to a situation where there is no sanctioning option. Hence, we observe net crowding out only in case of a revealed sanction decision and not in case the decision is hidden. The ex-ante commitment to sanction seems to entail a signal that causes the crowding out. This speaks in favor of agents interpreting the principal's ex-ante commitment to sanctions as a signal of distrust and lowering their amount of voluntary cooperation in response. Even if agents believe they will be sanctioned in case of low performance, hidden sanctions carry no signal such that there is no negative crowding out effect, but only a positive incentive effect.

The majority of agents react negatively to a revealed policy of sanctioning low performance. Apparently, those agents who negatively react to a revealed sanction policy, follow

the "presumption of innocence" concerning the principal's hidden decision. That is, these agents do not show any reaction to the hidden incentive even though they predominantly believe that the principal chose to sanction low performance. There are also agents who react positively on revealed sanctions (revealed as well as hidden): These agents increase their performance level in case of a revealed sanction decision in order to avoid being sanctioned. When sanctions are hidden, most of these subjects choose the same performance level as in the case where they definitely know that low performance will be sanctioned. These agents firmly believe that principals will sanction low performance and choose their performance level accordingly. In sum, we observe a slightly positive reaction in case of hidden sanctions as compared to a situation with no sanctioning option and an overall negative effect in case of a revealed sanction policy. In the revealed sanction case, committing to sanctions provokes strong crowding-out effects that outweigh the potential benefits.

The remainder of this article is organized as follows. Section 2 relates our study to the existing literature and identifies its contribution. Section 3 presents the design of the experiment and provides the experimental procedures and theoretical predictions. Results are reported in section 4. Section 5 concludes.

2 Related research

Starting point for our analysis is the literature on the detrimental effects of sanctions (e.g. Fehr and Rockenbach, 2003; Houser et al., 2008, Dickinson and Villeval, 2008). While Fehr and Rockenbach (2003) as well as Houser et al. (2008) analyze the effects of sanctions in the context of a modified investment game, Dickinson and Villeval (2008) use a principal agent framework with monitoring in order to study the effects of sanctions in the form of wage deductions. In all cases, the investor or the principal respectively *ex ante* commits to her sanctioning or monitoring policy, and in all cases, an ex ante commitment to impose sanctions or to choose a high monitoring intensity shows to be detrimental: Trustees reduce their back-transfers and agents choose lower effort levels when investors decide for the fining option or for a high monitoring intensity respectively (the latter leading to a larger probability of detecting and sanctioning low performance). While all three studies analyze the effects of a revealed ex-ante decision to sanction as opposed to a a revealed ex-ante decision not to sanction (Houser et al, 2008 further add a treatment with exogenous sanctioning), none of the existing studies so far compared ex-ante and ex-post sanctioning.

This is also true for the study of Falk and Kosfeld (2006) that serves as a major reference point for our analysis. Falk and Kosfeld do not focus on sanctions, but on the effects of control. In their experimental principal agent framework, the principal can either choose to control the agent by implementing a minimum performance requirement, hence restricting the agent's choice set, or she can leave the effort decision to the full discretion of the agent. After the principal has decided to either restrict the agent's choice set or not, the agent chooses a productive activity x that is costly for him $(-x)$ but beneficial for the principal $(+2x)$. Unlike the principal, the agent is provided with an initial endowment (120). In case the principal chooses to restrict the agent's choice set by imposing a predefined minimum requirement $\underline{x}$, x cannot fall below that requirement; in case she leaves the choice set unrestricted, the lowest possible choice of x is zero. To elicit agent choice, Falk and Kosfeld

(2006) apply the strategy method, i.e. they ask agents to choose x, both in case the principal decides to restrict the choice set (control) or in case she does not (trust). In their experimental analysis, Falk and Kosfeld (2006) vary the size of the predefined minimum requirement $\underline{x}$ leading to three different treatments (low, medium and high control).

The basic result of Falk and Kosfeld (2006) is that control has a negative effect on agent performance. While the "hidden costs of control" outweigh the benefits in all three control treatments, it is evident that even further raising the level of control (by increasing the minimum level of performance $\underline{x}$) will eventually lead to a situation where the direct benefits of control outweigh its (hidden) costs. Concerning different agent types, Falk and Kosfeld also observe agents who react positively or neutrally to the principal's decision to control, but those who react negatively, i.e. who reduce their performance level in light of control, represent the majority in all three treatments.

Our experiment is designed to supplement the evidence by Falk and Kosfeld (2006) in *three* important ways. *Firstly*, we change the type of control the principal disposes of. While in the study undertaken by Falk and Kosfeld the principal can actually restrict the agent's choice set and enforce a predefined minimum performance level, in our study the principal can only impose a fine in case the agent performs below a predefined minimum performance level. Hence, in our experiment the agent may well choose to perform below the required performance level - something which is precluded in the experiment run by Falk and Kosfeld. By assuming enforceable choice restriction by principals, Falk and Kosfeld implemented a new and so far unexplored option in experimental principal-agent settings. Our design tests whether the effects detected by Falk and Kosfeld also hold within a more conventional setting of presumably high practical relevance: a situation where the principal may not actually restrict the agent's choice set, but where she may use punishment as an enforcement device. *Secondly* and most importantly, our experiment adds a new and so far unexplored reference treatment to the analysis: hidden sanctions. While Falk and Kosfeld show that the revealed ex-ante decision to control decreases performance, their study remains silent on the effects of ex-post punishment. By comparing revealed and hidden sanctions within one experimental design, we are the first to shed light on the question if it is the *ex ante* commitment to sanction that drives the detrimental effects of sanctions or if sanctions as such are detrimental to performance. *Thirdly*, other than Falk and Kosfeld, we use a within-subject design and are hence able to compare the effects of different sanction schemes on subject level and to distinguish between different agent types as far as their differential reactions to the analyzed sanction schemes are concerned.

3 Experiment design, procedures and theoretical predictions

3.1 Experimental design

We employ a simple two stage principal-agent game relying on the parameter values in the medium control treatment (C10) of the experiment run by Falk and Kosfeld (2006). Agents receive an initial endowment of 120 and choose an activity level $y \in \{0, 1, ..., 120\}$, which is costly to the agent, but beneficial to the principal. The cost for the agent is $c(y) = y$ and

the benefit to the principal is $2y$, i.e. the marginal cost of choosing a higher activity level is always lower than the marginal benefit. The payoff functions are thus given by $\pi_R = 120 - y$ for the agent and $\pi_R = 2y$ for the principal.

We run three different treatments:

- In treatment REV-S (*revealed sanction*), principals have the option to fine the agent. They can either impose a fine $f = 10$ in case the activity level y of their agent does not exceed $\underline{y} = 10$, or they can refrain from doing so. If sanctioned by their principal, agents offering $y \leq \underline{y}$, receive a payoff of $\pi_R - f$. When offering $y > \underline{y}$, agents are released from the fine. If not sanctioned by their principal, agents are not exposed to any payoff reductions - irrespective of the activity level chosen.

- In treatment HID-S (*hidden sanction*), principals may also decide to sanction the agent in case of low performance ($y \leq \underline{y}$), but this decision is hidden from the agents such that these have to choose their activity level without knowing whether the principal chose to sanction low performance or not (but they know she may have chosen to sanction low performance). The payoffs are as in the revealed sanction treatment: Agents choosing $y \leq \underline{y}$ that are sanctioned by their principal, receive a payoff of $\pi_R - f$. If they are not sanctioned they are released from the fine. The same is true if they choose $y > \underline{y}$.

- In a reference treatment No-S (*no sanction*), there is no sanction option. Agents can choose their activity level without facing any fines. This treatment does not require a decision by the principal.

In order to elicit agents' choice of activity level in the revealed sanction treatment (REV-S) we made use of the strategy method. That is, before knowing the principal's sanctioning decision, agents stated their choice of activity level for both possible cases: (1) the case where the principal decided to impose a fine, and (2) the case where the principal decided to refrain from the fine.[1] In the hidden sanction treatment (HID-S), on the contrary, agents cannot condition their choice of activity level on the principals' sanctioning decision. They hence have to choose their activity level y independently from the principal's sanctioning decision. The same is true for the principal: She has to decide either to sanction or not in case of ($y \leq \underline{y}$), without knowing the agent's choice of activity level.

3.2 Experimental procedures

The experiment was conducted in the experimental laboratory of the Max Planck Institute of Economics in Jena, Germany. Subjects were students of the Friedrich-Schiller University of Jena, and were recruited using the online recruitment system ORSEE (Greiner 2004).

Subjects were randomly and anonymously paired, and their identities were never revealed to one another. At the beginning of the experiment, nature randomly determined the identity of the principal and the agent in each pair. Subjects then kept their role during the whole experiment. Each subject played one round of all three treatments in the same role. To

[1] In the experiment run by Falk and Kosfeld (2006) the results gained from the strategy method were validated by running an additional control treatment applying the specific response method, and it was shown that results were not an artefact of the strategy method.

Session	n	Treatment order
1	32	REV-S, HID-S, NO-S
2	32	HID-S, REV-S, NO-S
3	32	NO-S, REV-S, HID-S
4	32	NO-S, HID-S, REV-S

Table 1: Treatment orders of all experimental sessions

prevent learning effects, partners' decisions were only revealed after all decisions for all treatments had been made. Treatment order was altered within different sessions to test for order effects. Table 1 depicts treatments and their orders of each session as well as the number of participants. Each session was conducted with 16 agents and 16 principals, for a total of 128 subjects. Upon entering the laboratory, subjects were randomly assigned to 32 visually isolated computer terminals. Instructions for all required treatments (reproduced in the Appendix A) were distributed and read out loud by the experimenter. Questions were answered individually at the subjects' seats. Before beginning with the experiment, subjects filled out a short questionnaire testing comprehension of the rules for the three treatments. Each point earned in the experiment was exchanged for €0.1. Subjects earned on average €6.91 and received additionally a fixed participation fee of €2.50. Subjects were also asked to state their first-order beliefs concerning their partner's choices after each treatment.[2] Agents where asked:

Question 1: Do you believe your partner chose to sanction you with $f = 10$ in case you transferred $y \leq 10$?

Question 2: What do you think: Out of 100 participants, how many (principals) chose to sanction their (agents) and how many decided not to sanction their (agents)?[3]

Principals were asked the following two questions:

Question 3: How many points do you believe the (agent) has transferred to you?

Question 4: How sure are you with your guess (on a scale of 1 to 5)?

In the revealed sanction treatment (REV-S), principals had to answer question 3 twice: (a) for the case of an imposed sanction and (b) for the case that no sanction was imposed.

At the end of the three treatments, subjects received feedback on their partner's decisions in the three treatments and the resulting payoffs. Only one treatment was randomly chosen for payment. The experiment was programmed and conducted with z-Tree (Fischbacher 2007).

[2]Answers to these questions were not incentivized. There is evidence that eliciting expectations with or without monetary rewards does not yield significantly different results (see Grether, 1992; Camerer and Hogarth, 1999).

[3]We used neutral language to describe the two roles of the experiment: Principals and agents were called subjects A and B.

3.3 Theoretical predictions

The behavioral predictions for our experiment depend on the assumptions concerning subjects' preferences.

selfishness: When selfish players are assumed, the game's outcome remains constant irrespective of the considered treatment: A payoff maximizing agent will always choose an activity level of $y = 0$ irrespective of the principal's sanctioning decision.

inequity aversion: There is a ample evidence (from the laboratory as well as from the field) that many people do in fact not behave in a purely selfish manner. Recent theories try to capture these findings by assuming that people have a concern for equitable outcomes (see e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). In our experiment one might hence expect a substantial fraction of agents to choose an activity level $y > 0$. In light of the rather low minimum required activity level $\underline{y} = 10$ representing only 8.3 percent of the agent's initial endowment, one might hence expect inequity averse agents - much as the purely selfish ones - not to be affected by a sanction being imposed in case of low performance (neither hidden nor revealed). Instead, substantially inequity averse agents will always choose the same activity level $y > \underline{y}$ irrespective of the treatment. Only agents with a very low degree of inequity aversion (for an analogous argumentation see Falk and Kosfeld, 2006) may react to the fact of being sanctioned in REV-S and increase their activity level up to $\underline{y} + 1$. In the hidden sanction treatment (HID-S) their decision will also depend on their beliefs concerning the principal's sanctioning decision. If they firmly believe to be sanctioned in case of low performance they will choose the same activity level than in the REV-S case where the principal chose the sanctioning option and they definitely know to be sanctioned.

crowding out: As suggested by findings such as as those generated by Falk and Kosfeld, it may well be that an agent's intrinsic motivation for voluntary cooperation is "crowded out" by the installment of extrinsic incentives (rewards as well as sanctions or acts of control). Originally building on cognitive evaluation theory (Deci, 1971; Deci et al., 1989) crowding out is increasingly recognized as an important concept in economics. Following cognitive evaluation theory, extrinsic incentives may reduce an agent's intrinsic motivation to perform a given task by reducing his self-determination and self-esteem (Frey, 2008). The agent may hence interpret extrinsic incentives as a signal of distrust (Dickinson and Villeval, 2008; Falk and Kosfeld 2006) or as a signal of a low work norm (Sliwka, 2007) and lower his amount of voluntary cooperation in response.

In the context of our experiment, the ex-ante revealed decision to sanction low levels of activity might carry a *signal of distrust* that leads intrinsically motivated agents to reduce their activity level. These agents react to the negative signal by choosing a lower performance than they choose if the principal has no option to sanction the agent (crowding out effect). Only for those agents with a very low initial level of intrinsic motivation, the positive incentive effect of the sanction might outweigh the crowding out effect. Hence for these agents, deciding to sanction low performance and revealing this decision to the agent may lead to a positive reaction. On the other hand, the active decision not to sanction may also carry a *signal of*

trust. That is, even though allowed to sanction, the principal deliberately waives this option and this positive signal leads agents to increase their activity levels.

Finally, in the hidden sanction scheme where the agent does not know if he will be sanctioned or not in case of low performance, there is *no signal* conveyed. Hence, we can distinguish whether it is the act of sanctioning as such that crowds out intrinsic motivation or the signal-carrying threat associated with the ex-ante announcement. Concerning different types of agents, we would expect intrinsically motivated agents not to reduce their activity level as compared to a situation where there is no sanction option. This will be the case irrespective of whether the agent believes to be sanctioned or not, as it is not the act of sanctioning as such that crowds out his intrinsic motivation, but the signal-carrying threat associated with the ex-ante announcement. In the absence of crowding out for the intrinsically motivated, the activity levels chosen in the hidden sanction treatment might then on average be larger than the ones chosen in the no sanction treatment. The reason is that agents with a negligible initial level of intrinsic motivation will presumably act according to their believes and - whenever they expect to be sanctioned in case of low performance - will choose higher activity levels as compared to the no sanction situation.

4 Results

Concerning our experimental results, we first report on agents' behavior on the aggregate and individual level - explicitly comparing our results with those by Falk and Kosfeld. In a next step, we discuss the principals' decisions and beliefs.

4.1 Agents' behavior

Table 2 shows agents' average and median performance in the REV-S, HID-S, and NO-S treatment. For the REV-S treatment, the table reports both, the agents' activity levels in case the principal decides to sanction the agent and in case she decides to refrain from the sanctioning option. As the average and medium activity levels chosen by the agents are substantially larger than zero, agents on average are apparently not characterized by pure selfishness. But also inequity aversion apparently cannot explain the full range of behavior found.

Concerning potential crowding out effects, our first result is based on the comparison of the revealed sanction (REV-S) and the hidden sanction (HID-S) treatment (each in relation to the no-sanction (NO-S) case):

RESULT 1: *We only observe net crowding-out in REV-S, but not in HID-S: In REV-S, average agent performance is lower if the principal uses the sanctioning option as compared to the No-S case. In HID-S, average agent performance is higher than in NO-S. If the principal refrains from sanctioning in REV-S, average agent performance is not significantly different from No-S.*

Support for Result 1 comes from Table 2: The average and median transfer choices in the REV-S treatment show that a revealed sanction entails hidden costs: Agents choose

significantly lower activity levels in face of a revealed sanctioning decision (REV-S) than in the NO-S treatment (Wilcoxon signed rank test, $p<0.0297$; paired t-test, $p<0.0077$). The difference is even more pronounced when we compare the average transfer choices in the REV-S treatment when the principal either decided in favor of a sanction policy or against it: Agents' average performance is significantly less when principals decided to use the sanction option than when they decided not to do so (Wilcoxon signed rank test, $p<0.0027$; paired t-test, $p<0.0011$). In accordance with Falk and Kosfeld, we hence observe that deciding in favor of sanctioning or controlling entails "hidden costs" as compared to actively deciding against it: Agents may indeed interpret active and pre-announced sanctions and acts of control as a *signal of distrust* and therefore lower their performance.

Furthermore, agents might also interpret the principal's active decision against sanctioning as positive *signal of trust*: As Table 2 reveals, average transfer choices in case of a deliberate no-sanction policy in REV-S are higher than those when there is no sanction option available (No-S). However, the displayed differences in transfer choice between opting against sanctioning in REV-S and No-S are not statistically significant (Wilcoxon ranked sum test, $p>0.2552$; paired t-test, $p>0.2767$). I.e., actively choosing to sanction and revealing this decision to agents entails hidden costs, but opting against sanctioning and revealing this decision entails no or only marginal hidden benefits.

But what about hidden sanctions? The comparison of agents' average transfer choice in REV-S where the principal chose to sanction low performance with those in HID-S, reveals that agents clearly discriminate between both sanction policies: Agents' average transfer choices are much smaller in case of a revealed sanction choice than in case they do not know whether the principal decided to sanction low performance or not (Wilcoxon signed rank test, $p<0.0001$; paired t-test, $p<0.0001$). While the use of the sanction decreases work compliance in the revealed sanction case, the contrary holds for the hidden sanction case: Agents transfer significantly more in HID-S than they do in NO-S (Wilcoxon signed rank test, $p<0.0343$; paired t-test, $p<0.0459$). This is in spite the fact that almost all agents (90 percent) in HID-S believe they will be sanctioned in case of low performance. Hence, although agents believe to be confronted with the same sanction choice as in the revealed sanction case where they definitely know to be sanctioned (REV-S), they behave differently in HID-S. This suggests that there is a fundamental difference between both sanction policies: While the active decision to sanction in the REV-S treatment seemingly carries a signal of distrust and results in net crowding out, there is no such signal in HID-S. In HID-S, the agent does not know whether the principal opted in favor or against sanctioning low performance. The lack of such a signal of distrust apparently prevents crowding out. At the same time, however, the fact that agents predominantly believe to be sanctioned in case of low performance in HID-S has positive incentive effects. As a result, average activity levels in HID-S are significantly larger than those in NO-S.

This suggests: The use of the sanction only entails hidden net costs in the revealed case. When the decision to use the sanction is hidden, we find a positive net effect of the sanctioning option. Hence, net crowding-out appears to be limited to revealed sanction decisions.

Result 1 is further highlighted and supported by figure 3 depicting the cumulative distributions of activity levels in the different treatments. For reasons of clarity we present pair-wise comparisons: Figure 1A compares the cumulative distribution of activity levels

	REV-S		HID-S	NO-S
	Sanction	No sanction		
Number of participants	64	64	64	64
Average activity level	16.2	19.6	20.3	18.8
Median activity level	11.5	17.5	15	15

Table 2: Agents' average and median activity levels in the REV-S, HID-S, and NO-S treatment.

in the REV-S treatment in case the principal chose to sanction (grey line) and in case he refrained from doing so (black line). Figure 1B compares REV-S activity levels in case the principal chose to sanction (grey line) to HID-S activity levels (back line). Finally, figure 1C compares HID-S activity levels black line) to NO-S activity levels (grey line).

Figure 1C reveals that the positive effect of the hidden sanction as compared to the NO-sanction case is indeed driven by an incentive effect as it stems from a shift of activity level choices from $0 \leq y \leq \underline{y}$ to activity levels of $y = 11$: In HID-S, only 22 percent of the agents choose an activity level in the range of $0 \leq y \leq \underline{y}$, but 38 percent do so in the NO-S treatment, and 17 percent choose an activity level of exactly $y = 11$ in HID-S as compared to only 5 percent in NO-S. Agents apparently increase their activity levels just above $y = 10$ to avoid being sanctioned (incentive effect). The cumulative distributions reveal that while for each value of $0 \leq y < 11$, there are always strictly more agents who choose at least the same activity level y in HID-S than in the NO-S treatment, above $y = 11$ there are no differences between the treatments. I.e., the on average higher transfer levels in HID-S as compared to No-S are essentially driven by the incentive effect for those agents who would have otherwise preferred to choose a lower activity level. The fact that agents in HID-S in fact predominantly believe to be sanctioned in case of low performance underlines this presumption.

Concerning a comparison between HID-S and REV-S in case the principal opted to sanction low performance (figure 1B), we observe an overall shift in the distribution: For each activity level y, there are always strictly more agents who choose at least y in HID-S than there are in the REV-S treatment where principals chose to sanction. For example, there are only 22 percent of agents who choose an activity level below $y = 11$ in HID-S, but 30 percent do so in REV-S. The shift in distributions emphasizes the different impact of the revealed and hidden sanction policy on agents' behavior. The figure shows that only revealed sanctions turn out to be costly and induce net crowding-out. At the same time, both sanction policies seem to have comparable positive incentive effects: They both increase the share of agents transferring exactly $y = 11$ (REV-S sanction: 20 percent, HID-S: 17 percent, both in comparison to REV-S no sanction: 2 percent and NO-S: 5 percent).

Figure 1A finally demonstrates that the principal's decision to sanction indeed crowds out agents intrinsic motivation. Here we compare the cumulative distributions of REV-S transfers in case the principal decides to sanction low performance and in case he does not. The crowding out effect of the sanction decision emerges in the area of activity levels over and above $y = 11$: For each value of $y \geq 11$, there are always strictly more agents who choose at least the value of y in REV-S no sanction than in the REV-S sanction case. For example, there are only 50 percent of agents who choose an activity level over and above

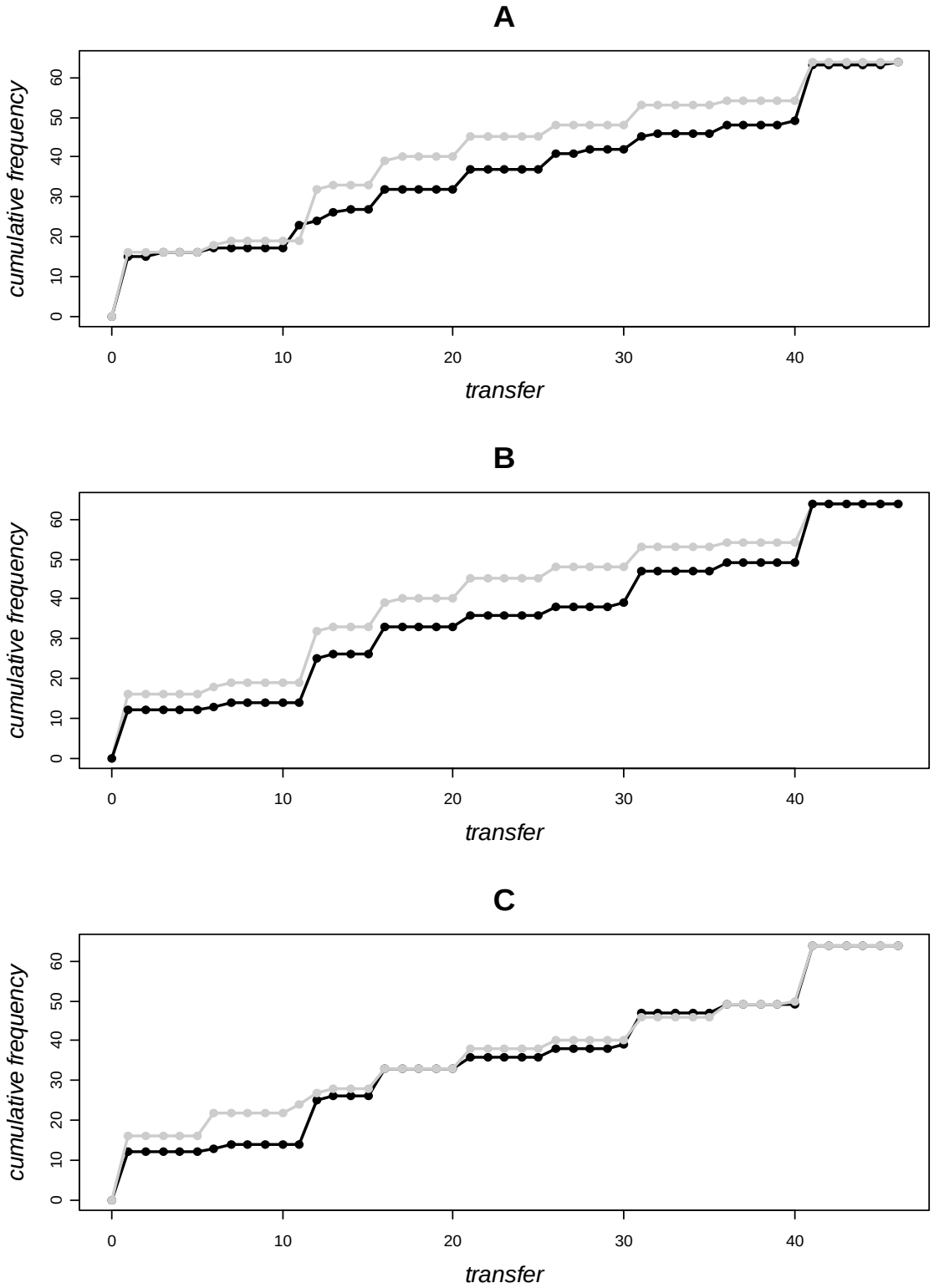


Table 3: Cumulative distribution of agents' activity levels (A: REV-S- sanction (gray) vs. REV-S- no sanction (black), B: REV-S - sanction (gray) vs. HID-S (black), C: HID-S (black) vs. NO-S (gray))

$y = 11$ in the REV-S sanction case, but 62 percent do so in the REV-S no sanction case. Comparable to the study by Falk and Kosfeld on the hidden costs of control, we find hidden costs of revealed sanctions. Other than the experiment run by Falk and Kosfeld, our design does not preclude agents from choosing activity levels $y \leq 10$. We find that below $y = 11$, there are no differences in distributions between a revealed policy to either sanction or not. Potentially in that range, incentive and crowding out effects cancel out.

This suggests: Positive effects of hidden sanction as compared to the no-sanction treatment emerge (a) because low performing agents increase their activity level to avoid the sanction (incentive effect) and (b) there is no crowding out as the hidden decision to sanction or not - by definition - cannot carry a signal (figure 1C). While the revealed sanction policy entails positive incentive effects as well, these are - much as in the experiment run by Falk and Kosfeld - over-compensated by their hidden costs (figure 1A). These costs presumably result from a revealed sanctioning policy being interpreted as a signal of distrust leading to the crowding out of intrinsic motivation. Hidden costs of sanctions are hence exclusively connected to revealed sanction decisions - a fact which is further highlighted by directly comparing the situation of a revealed sanction policy with a hidden sanction policy (figure 1B).

In what follows we take a closer look at individual agents' behavior. Since we asked agents to choose activity levels for all three treatments, we can distinguish between different *types of agents* concerning their differential behavioral patterns in the different treatments. Since we used the strategy method in the REV-S treatment, we can distinguish agents by their reaction to the principal's sanctioning decision. They might either respond positively, neutrally or negatively to that decision. We find:

RESULT 2: *There is a strong heterogeneity among agents. We observe agents who react positively, neutrally, or negatively to the revealed principals sanction decision. The last group is much larger than the first.*

Support for Result 2 comes from table 4: While 16 percent of agents react positively to the principal's decision to sanction in the REV-S treatment, a significantly larger share of agents, namely 42 percent, react negatively (chi-squared, $p>0.0018$). The results resemble very much the ones derived by Falk and Kosfeld[4]. However, the share of positively reacting agents in our experiment is even smaller than the one in the experiment run by Falk and Kosfeld where 25 percent of agents react positively to the revealed implementation of control. The difference, however, is easily explained by the fact that in the experiment run by Falk and Kosfeld agents are literally forced to at least fulfill the minimum performance requirement whereas in our design they may also decide for low performance and pay the corresponding fine. Without perfect enforcement, agents may deliberately choose not to positively respond to the incentive. Clearly, this liberation implies a smaller fraction of positively responding agents. Interestingly, however, we also find less agents that react negatively to the revealed decision to sanction (42 percent) than Falk and Kosfeld do to the revealed implementation of control (57 percent). At first, this might seem surprising, because from the agents' point

[4]We find a significantly larger share of agents who do not respond ("neutral reaction") to the incentive, chi-squared, $p<0.0038$

		Positive	Neutral	Negative
Klempt and Pull REV-S	Number of participants	10	27	27
	Relative share	0.16	0.42	0.42
	Average transfer if sanctioned	15.7	19.2	13.4
	Average transfer if not sanctioned	6.7	19.2	24.7
Falk and Kosfeld C10	Number of participants	18	13	41
	Relative share	0.25	0.18	0.57
	Average transfer if controlled	11.1	22.7	18.7
	Average transfer if not controlled	1.9	22.7	32.3

Table 4: Agents' reaction to principals' choices in different experiments

of view being sanctioned or controlled results in earning at least 110 in each of the two experiments: E.g., agents who choose an activity level $y = 0$ and are subsequently sanctioned in our experiment earn 110; agents who transfer the minimal requirement in the experiment run by Falk and Kosfeld also receive 110. Again, one might explain the difference in behavior by the lower degree of restrictiveness of the sanction incentive as compared to the implementation of control. The latter may rightfully be perceived as a stronger intervention than the implementation of a fine. Consequently, the signal of distrust might be more profound.

More than just identifying different types of agents according to their differential reaction to the implementation of sanctions in the revealed sanction treatment (REV-S), our within-subject design also allows us to analyze how the identified types of agents act in the hidden sanction (HID-S) or no sanction (NO-S) treatment. We observe:

RESULT 3: *Agents who react negatively to the revealed sanction decision (REV-S) do not react at all to the hidden sanction (HID-S) as compared to the no sanction situation (No-S). Agents who positively react to the revealed decision (REV-S), and believe that they also will be sanctioned in the hidden case, react positively to the hidden decision (HID-S) as compared to a no sanction situation (NO-S).*

Support for Result 3 comes from Table 5. This table first repeats the absolute and relative frequency of agents who react positively, neutrally, or negatively to the revealed principal's decision to sanction in REV-S and further displays the average transfer choices of these agent types in all three treatments (REV-S, HID-S and NO-S).

Interestingly, the three agent types (positive, neutral, negative reaction to sanction in the revealed sanction case as compared to no sanction in the revealed sanction case) do not only show different behavioral patterns when confronted with a sanction or no-sanction policy in the REV-S treatment, but they also show interesting behavioral patterns in the HID-S and in the NO-S treatment.

		Positive	Neutral	Negative
Number of participants		10	27	27
Relative share		0.16	0.42	0.42
REV-S	Average transfer if sanctioned	15.7	19.2	13.4
	Average transfer if not sanctioned	6.7	19.2	24.7
HID-S	Average transfer	14	21.3	21.6
NO-S	Average transfer	10.6	19.1	21.6

Table 5: Agents' types dependent on their reaction to the revealed principal's decision to sanction and their choice of activity level

We first take a closer look at those agents who react *negatively* to a sanction policy in the REV-S treatment. Their average activity level in REV-S goes back from $y = 24.7$ in case the principal decides against sanctioning to only $y = 13.4$ in case she opts to sanction low performance - a truly substantial decline. When there is no sanctioning option (NO-S) these very same agents on average choose an activity level $y = 21.6$. Both differences are statistically significant (Wilcoxon ranked sum test [Paired t-test], $p<0.0003$ [$p<0.0001$] for the difference in average activity levels between No-S and a revealed sanction policy in REV-S and $p<0.0087$ [$p<0.0068$] for the difference in average activity levels between No-S and a revealed no-sanction policy in REV-S). Apparently then, for those agents that react negatively to a sanction policy in REV-S, both, the decision to sanction as well as the decision not to do so represent a signal: While the former is perceived as a signal of distrust leading to lower activity levels as compared to a situation where there is no sanction option (crowding out of intrinsic motivation), the latter may serve as a signal of trust leading to higher levels of activity (further enhancing the initial level of intrinsic motivation, so-called "crowding in"). Hence, unlike Falk and Kosfeld we find evidence for agents punishing the principal's decision to sanction or control as well as evidence for agents rewarding the principal's decision to actively waive his sanction opportunity. Strikingly, agents who negatively react to the principal's revealed decision to sanction show no behavioral disparity whatsoever between the hidden sanction (HID-S) and the no sanction (NO-S) treatment. In both cases they choose the same level of activity ($y = 21.6$; Wilcoxon ranked sum test [Paired t-test], $p>0.9254$ [$p>0.9801$]). This is the case even though 90 percent of these agents believe that the principal decided for a sanction policy. Apparently, they follow the "presumption of innocence" concerning the principal's sanction decision. As they do not observe the principals' decision they neither want to punish nor reward that decision.

How about agents who *positively* react to the sanction option in the REV-S treatment? Their average activity level in REV-S increases from only $y = 6.7$ in case the principal decides against sanctioning to as much as $y = 15.7$ in case she opts to sanction low performance. For these agents whose activity level is mainly driven by external rewards, the incentive effect of sanctioning apparently overcompensates any potential crowding effect. Interestingly, these agents even choose a lower activity level when the principal actively waives the option

to sanction in the revealed sanction case (REV S) than in the no-sanction case (NO-S). Apparently these basically extrinsically motivated agents feel that the principal "deserves" to be left with a lower amount when she actively gives away the only measure of influence she disposes of. When sanctions are hidden (HID-S), these very agents - as one might expect - act according to their believes: Those who believe to be sanctioned in the HID-S case (90 percent) choose the same activity level than in the sanction case in REV-S and significantly more than in NO-S (Wilcoxon ranked sum test [Paired t-test], $p>0.3173$ [$p>0.3466$] for the first comparison and $p<0.0316$ [$p<0.0232$] for the latter, positively reacting agents who believe to be sanctioned in HID-S state the following average transfer choices: REV-S sanction: 16.2, REV-S no sanction: 7.4, HID-S: 15.6 and NO-S: 11.8). As there is only one agent in the class of positively reacting agents who does dot believe to be sanctioned in HID-S, we refrain from interpreting the respective results.

To conclude, Table 5 sheds light on the reason why crowding-out effects disappear when sanctions are hidden: On the one hand, those agents who provoke the crowding-out effect in the revealed sanction case by decreasing their work compliance do not react to the hidden sanction option. And on the other hand, the positive effect of the sanction remains to be in place as those agents who positively react to the revealed sanction decision in REV-S, also do so in the hidden sanction treatment.

4.2 Principals' behavior

We now report the behavior of the principals. Given the average agents' choices, principals should not choose to sanction the agents in REV-S. In HID-S, the principals' decision does not influence agent choices, and principals do not have to consider signaling effects. But in spite of that, we observe the following:

RESULT 4: *The majority of principals chooses to sanction the agents in REV-S and in HID-S, and agents correctly anticipate that behavior.*

We find that 73 percent of principals choose to sanction the agents in REV-S. In HID-S, even a higher percentage of principals (84 percent) chooses to sanction. This difference is weakly significant (chi-squared test, $p<0.09682$) and might hint at (some) principals fearing to send a signal when choosing to sanction in the revealed sanction case. Agents correctly anticipate principals' sanctioning behavior: In REV-S, 80 percent of the agents ex-ante believe that the principal will decide to sanction (ex-post they are perfectly informed), and 90 percent do believe so in HID-S (chi-squared test shows that both beliefs are correct, $p>0.42$ for both comparisons). Again, this result might be interpreted in the sense that (some) agents expect principals to refrain from sending a signal of distrust by actively deciding in favor of a pre-announced sanction option.

But why do so many principals sanction in REV-S although it is not payoff maximizing? One might explain the high fraction of principals who sanction in REV-S with their inability to predict the full hidden costs it entails. A comparisons of principals' beliefs and agents' actual choices of activity level in table 6 supports this presumption. Strikingly, principals expect agents to choose lower activity levels than these actually do in all three treatments. In case of a sanction decision in REV-S, they expect to receive 5.6 less on average than is actually

	REV-S		HID-S	NO-S
	Sanction	No sanction		
Number of participants	64	64	64	64
Agents'average choice of activity level	16.2	19.6	20.3	18.8
Principals' belief	10.6	10.3	14.4	7.9

Table 6: Agents' choice of activity levels and principals' believes concerning agent choice.

transferred. This difference even adds up to 10.9 in NO-S. All differences are statistically significant (Wilcoxon rank sum test [t-test], $p<0.01344$ [$p<0.0331$] for all comparisons). Further, Table 6 also reveals that principals do not expect to receive different transfers in REV-S when either choosing to sanction or choosing not to do so (Wilcocon signed rank test [paired t-test], $p>0.2550$ [$p>0.8464$]). On average, they expect to receive 10.6 when sanctioning and 10.3 when not. In both cases they expect to receive even less in the no-sanction (NO-S) case. On average, they expect their active decision not to sanction in REV-S to carry a signal of trust and hence increase agents' activity level as compared to the NO-S case. But in case they decide to sanction they expect potentially negative crowding-out effects to be leveled out by positive incentives effects. On average, they clearly underestimate the negative crowding out effects of their sanction decision leading them to choose the sanctioning option in REV-S. Concerning the effect of the hidden sanction option, principals correctly predict it to be positive: They expect to receive 14.4 in HID-S as compared to 7.9 in NO-S, but overestimate its size (the predicted difference between both treatments is significantly larger than the actual difference; Wilcoxon signed rank test [paired t-test], $p<0.0059$ [$p<0.0115$]).

Our finding that the majority of principals chooses to sanction is in sharp contrast to the respective finding by Falk and Kosfeld, 2006. In their experiment, the majority of principals anticipates the negative effects of control and refrains from using the controlling option. Having asked principals for their beliefs concerning agent behavior in REV-S in case they use the sanctioning option or they deliberately waive it, we can shed light on this question by distinguishing different types of principals according to their belief on agents' reaction to the use of the sanction option in REV-S. We observe the following:

RESULT 5: *The majority of principals believes that agents will positively react to the revealed sanction. Those who believe in a positive reaction sanction more often in REV-S than those who believe that agents will negatively react to the sanction. In HID-S, there are no differences in the sanctioning behavior with respect to the principal's type concerning his beliefs.*

Support for result 5 is given by table 7. Table 7 depicts how many principals believe that agents will react positively, neutrally, or negatively to the use of the sanctioning option in REV-S and then displays principals' average beliefs on the choice of activity level for all the different treatments. The table also shows the fraction of principals who sanction in REV-S and HID-S - again dependent on their belief concerning the agents' reaction to a sanction policy in REV-S.

First of all, the table shows that more principals believe agents' to react positively than negatively to the sanction: 42 percent of the principals believe that agents will positively

		Positive	Neutral	Negative
Number of participants		27	16	21
Relative share		0.42	0.25	0.33
REV-S	Average transfer belief in case of sanction	13.4	5.9	10.5
	Average transfer belief in case of no sanction	6.7	5.9	23.2
	Fraction of principals who sanction	0.85	0.75	0.57
HID-S	Average transfer belief	14.4	6.6	20.2
	Fraction of principals who sanction	0.88	0.81	0.81
NO-S	Average transfer belief	4.5	4.8	14.6

Table 7: Principals' beliefs concerning agents' reaction to the sanction

react to the sanction (net incentive effect), whereas only 32 percent believe in the opposite (net crowding effect). This is in sharp contrast to real behavior found. As principals' sanctioning decisions in REV-S should depend on their belief concerning agent types, part of principals' sanctioning behavior might indeed be explained by their counter-factual believes concerning agents' reaction to being sanctioned or not. However, even those principals who believe agents to react negatively to the sanction in REV-S choose to sanction in 57 percent of the cases. These principals sanction although they believe that it will make them worse off. This hints at other underlying concerns motivating principals' sanctioning behavior beyond mere payoff maximization. Accordingly, those principals who expect agents to positively react to a sanction but still refrain from doing so (15 percent), might derive some unexplored benefit from actively waiving their sanction option in REV-S besides being interested in their own monetary payoff.

Concerning the HID-S treatment, we do not observe any differences in principals' sanctioning behavior with respect to their belief type: between 81 and 88 percent of principals decide in favor of using the sanction option (chi-squared test, $p>0.7183$ for all comparisons). In comparison with the REV-S case, this result is mainly driven by those principals who believe in a negative agent reaction: When the sanction policy is hidden, there is no need to act according to an expected agent reaction.

4.3 Conclusion

In our paper, we analyze the differing impacts of hidden and revealed sanctions on agent performance in a simple principal-agent game. We are able to show that sanctions only entail net hidden costs if the decision to sanction is ex-ante revealed to the agent. If the sanction policy is not revealed to the agent such that he does not know whether he will be sanctioned or not, the fact that he might be sanctioned in case of low performance has a positive effect on agent performance (as compared to a situation where there is no sanctioning option).

The same is true if the principal actively waives his sanctioning option and this decision is communicated to the agent: Here, too, agent performance increases as compared to a situation where there is no sanctioning option in the first place.

Our results are compatible with and further enhance and support crowding out theory according to which a principal's ex-ante decision to sanction low performance might be interpreted by agents as a signal of distrust while the ex-ante decision to actively waive this option might be interpreted as a signal of trust (as compared to a situation where there is no sanctioning option at all). The former will tend to reduce ("crowd out") agents' intrinsic motivation, the latter will increase ("crowd in") agents' intrinsic motivation. Even if an ex-ante commitment to sanction will obviously not only have negative crowding out effects, but also positive incentive effects by leading otherwise low-performing agents to increase their activity level up to a point where they are not sanctioned, our experimental evidence clearly speaks in favor of net crowding out (crowding out effect outweighing incentive effect) in the revealed sanction case. As hidden sanctions, however, do not carry a signal and hence create no potential for crowding out effects, we here observe net incentive effects (incentive effects outweighing potential crowding out effects) with agents performing better in a situation of hidden sanctions than in a situation with no sanction option.

More than supporting and enhancing theory, our results also seem to mirror what is to be observed in real-life organizations: In many cases, real-life organizations do not rely on employment contracts specifying the details of work requirements and the sanctions to be imposed if these are not met. Among others, the absence of pre-announced sanctions in employment contracts could be motivated by the fear that explicit threats might offend workers and reduce their intrinsic motivation. At the same time, real-life organizations rely on a very severe ex-post sanctioning option by being able to fire a worker who has substantially fallen below the implicitly established work norms or not to renew his contract. The existence of this powerful ex-post sanctioning option has been shown to have substantial positive effects on workers' activity level (see e.g. the study with field data by Ichino and Riphahn 2005 or the experimental evidence delivered by Brown et al. 2001). These positive incentive effects of the firing option might arise because the opportunity to punish the agent ex-post disciplines low preforming agents without offending those who cooperate voluntarily with a pre-announced threat.

A Appendix: Instructions

Please read the following instructions carefully. The instructions will provide you with all the information you require for participation in the experiment. Please ask for assistance if there is something that you do not understand. Your question will be answered at your workplace. There is a strict prohibition of communication during the experiment.
You will receive an initial endowment of 2.50 euros at the beginning of the experiment. You can earn additional money over the course of the experiment by collecting points. All of the points which you accrue over the course of the experiment will be converted to euros at the end of the experiment. Please note that: 1 point = 10 cents.
At the end of the experiment, you will receive the income which you earned over the course of the experiment plus the 2.50 euros of initial endowment in cash.

The experiment

In this experiment, each participant A is associated with a participant B in a group of two. No participant knows with whom he is associated, meaning that all decisions are made anonymously. Participant A receives an amount of 120 points at the beginning of the experiment. Participant B receives no points.

Participant A's decision:

Participant A can decide how many points he wants to transfer to participant B. The experimenter doubles each point that participant A transfers to participant B. Thus, each point that participant A transfers to participant B reduces participant A's income by one point and increases participant B's income by two points.

The formula for calculating income is as follows:
Participant A's income: $120 - transfer$
Participant B's income: $0 + 2 \cdot transfer$

The following examples clarify the income formulas:
Example 1: A transfers 0 points to B. The incomes are then 120 for A and 0 for B.
Example 2: A transfers 20 points to B. The incomes are then 100 for A and 40 for B.
Example 3: A transfers 80 points to B. The incomes are then 40 for A and 160 for B.

Participant B's decision:

Participant B can decide to sanction participant A. That is, he can decide to reduce participant A's income by 10 points in case he transfers not more than 10 points. The deducted points only reduce the income of participant A and are not credited participant B. Participant B can also decide not to sanction participant A in case he transfers 10 or less.

Hence, there are two cases:

Case 1: Participant B decides to sanction participant A in case he transfers not more than 10 points. That is, participant A loses 10 points in case he transfers between 0 and 10 points. In case he transfers between 11 and 120 points, no points will be deducted from participant A's income besides the transfer.

Case 2: Participant B decides not to sanction participant A and allows participant A to freely decide on his transfer. In this case, participant A can transfer any (integer) amount between 0 and 120 to B without facing any additional income reductions.

Important: The 10 points deduction in case of a sanction decision are not credited to participant B's income but are only deducted from participant A's income.

The course of the experiment

The experiment consists of three parts - part 1, part 2 and part 3.
Either you keep the role as participant A in all three parts of the experiment and take a decision as described at "Participant A's decision". Or, you keep the role as participant B in all three parts of the experiment and take a decision as described at "Participant B's decision". First you take the decisions of part 1. Then, you take the decisions of part 2, and finally, you take the decisions of part 3.
You will learn the decisions of your counterpart as soon as both of you have made all the decisions for all three parts of the experiment. You will also only learn about your income at the end of the whole experiment. After all decisions have been made, one out of three parts will be picked randomly and only your earning gained in this part of the experiment will be paid out.

Part 1

Part 1 only consists of one stage. Only participant A decides.

Stage 1:
In stage 1, participant A decides which integer amount between 0 and 120 he will transfer to participant B.

Note: Participant B cannot decide to sanction participant A. Treatment 1 only requires a decision of participant A.
Part 1 of the experiment is over as soon as participant A decided how much to transfer to participant B. Participant B will be informed about participant A's decision at the end of the experiment.

Part 2

Part 2 consists of two stages that define the sequence of decisions.

Stage 1:
In stage 1, participant B decides whether or not he sanctions participant A in case participant A transfers less than than 10.

Stage 2:
In stage 2, participant A decides which integer amount between 0 and 120 he will transfer to participant B. It holds that:

- in case participant B chose to sanction participant A: If participant A transfers an amount between 0 and 10 to participant B, 10 points will be deducted from participant A's income. If participant A transfers an amount between 11 and 120 points, no points will be deducted from participant A's income, or
- in case participant B chose not to sanction participant A: No points will be deducted from participant A's income - irrespective of the size of the chosen transfer.

Participant A will not be informed about whether or not participant B chose to sanction a transfer of not more than 10 points. Participant A will only learn about participant B's decision of part 2 at the end of part 3. However, A can condition his transfer choice on participant B's decision in the first stage. Hence, participant A takes two decisions with the help of the following monitor:

(Here, the original instructions provide a screenshot)

Hence, participant A indicates how many points he will transfer to participant B in case participant B decided to sanction a transfer between 0 and 10 points (case 1) and in case he decided not to do so (case 2).
Which of the decisions is relevant depends on what participant B decided in the first stage. If participant B decided to sanction, the decision that participant A gave under case 1 applies. If participant B decided not to sanction, the decision that participant A indicated under case 2 applies.
Part 2 of the experiment is over as soon as participant A decided how much to transfer to participant B. You will be informed about the other participant's decision in part 2 at the end of the experiment.

Part 3

Part 3 consists of two stages that define the sequence of decisions.

Stage 1:
In stage 1, participant A decides which integer amount between 0 and 120 he transfers to participant B.

Stage 2:
In stage 2, participant B decides whether or not he sanctions participant A in case participant A transfers less than than 10. It holds that:

- in case participant B chooses to sanction participant A: If participant A transfers an amount between 0 and 10 to participant B, 10 points will be deducted from participant A's income. If participant A transfers an amount between 11 and 120 points, no points will be deducted from participant A's income, or
- in case participant B chooses not to sanction participant A: No points will be deducted from participant A's income - irrespective of his transfer choice.

Participant B will not be informed about how many points participant A transferred to him in the first stage. Participant B will only learn about participant A's decision at the end of treatment 3. Participant A takes his decisions of the first stage with the help of the following monitor:

(Here, the original instructions provide a screenshot)

Hence, participant A indicates how many points he will transfer to participant B.
Part 3 is over as soon as participant B decided whether he sanctions participant A or not. You will be informed about the other participant's decision in part 3 at the end of the experiment.

At the end of all three parts of the experiment a final monitor will inform you of the decisions and the resulting incomes. Your point income will be converted to euros and paid out in cash together with the initial endowment.

Do you have any questions?

References

Andreoni, J., W. Harbaugh, and L. Vesterlund (2003): "The Carrot or the Stick: Rewards, Punishments, and Cooperation," *American Economic Review*, 93(3), 893–902.

Bolton, G., and A. Ockenfels (2000): "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, pp. 166–193.

Camerer, C. (2003): *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press Princeton, NJ.

Camerer, C., and R. Hogarth (1999): "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of Risk and Uncertainty*, 19(1), 7–42.

Deci, E. (1971): "Effects of Externally Mediated Rewards on Intrinsic Motivation.," *Journal of Personality and Social Psychology*, 18(1), 105–115.

Deci, E., J. Connell, and R. Ryan (1989): "Self-determination in a work organization," *Journal of Applied Psychology*, 74(4), 580–590.

Dickinson, D., and M. Villeval (2008): "Does monitoring decrease work effort? The complementarity between agency and crowding-out theories," *Games and Economic Behavior*, 63(1), 56–76.

Falk, A., and M. Kosfeld (2006): "The Hidden Costs of Control," *American Economic Review*, 96(5), 1611–1630.

Fehr, E., and A. Falk (2002): "Psychological foundations of incentives," *European Economic Review*, 46(4), 687–724.

Fehr, E., and S. Gachter (2000): "Cooperation and punishment in public goods experiments," *American Economic Review*, 90(4), 980–994.

——— (2002): "Altruistic punishment in humans," *Nature*, 415(6868), 137–140.

Fehr, E., and B. Rockenbach (2003): "Detrimental effects of sanctions on human altruism," *Nature*, 422, 137–140.

Fehr, E., and K. Schmidt (1999): "A theory of fairness, competition, and cooperation*," *Quarterly journal of Economics*, 114(3), 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Frey, B. (1993): "Does monitoring increase work effort? The rivalry with trust and loyalty," *Economic Inquiry*, 31, 663–663.

Frey, B. S. (2008): "Motivation crowding theory - a new approach to behaviour," in *Behavioral Economics and Public Policy*, pp. 37–54. Australian Government Productivity Commission.

Gneezy, U. (2004): "The W effect of incentives," *Unpublished Paper*.

Gneezy, U., and A. Rustichini (2000a): "A Fine is a Price," *The Journal of Legal Studies*, 29(1), 1–17.

——— (2000b): "Pay Enough or Don't Pay at All*," *Quarterly Journal of Economics*, 115(3), 791–810.

Greiner, B. (2003): "An Online Recruitment System for Economic Experiments," *Forschung und wissenschaftliches Rechnen*, pp. 79–93.

Grether, D. (1992): "Testing Bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior & Organization*, 17(1), 31–57.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. Cardenas, M. Gurven, E. Gwako, N. Henrich, et al. (2006): "Costly punishment across human societies," *Science*, 312(5781), 1767.

Houser, D., E. Xiao, K. McCabe, and V. Smith (2008): "When punishment fails: Research on sanctions, intentions and non-cooperation," *Games and Economic Behavior*, 62(2), 509–532.

Ichino, A., and R. Riphahn (2005): "The effect of employment protection on worker effort: Absenteeism during and after probation," *Journal of the European Economic Association*, 3(1), 120–143.

Lazear, E. (2000): "Performance pay and productivity," *The American Economic Review*, 90(5), 1346–1361.

Sliwka, D. (2007): "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *American Economic Review*, 97(3), 999–1012.