

Can Sanctions Induce Pessimism? An Experiment*

Roberto Galbiati[†] Karl Schlag[‡] Joël van der Weele[§]

March 24, 2009

Abstract

We run an experiment in which two subjects play a two-round minimum effort game in the presence of a third player (principal) who is the only one informed about past effort choices and benefits from a higher minimum effort of the others. Sanctions introduced in the second round by the experimenter lead to more optimistic beliefs and higher efforts. This is not true when sanctions have been imposed by the principal. The possibility that the choice of a sanction is a signal of low effort levels causes players who chose high effort in the first round to be less optimistic.

Keywords: Sanctions, beliefs, expressive law, deterrence, coordination, minimum effort game.

JEL-codes: C92, D83, K42

1 Introduction

In the standard economic view, sanctions are effective because they change economic payoffs and modify individuals' incentives to engage in certain actions. More recently, the results from several experiments suggest that the effect of sanctions is not always so straightforward, and that sanctions can even be counterproductive in some situations (see Frey and Jegen (2001) and

*For valuable comments and discussions we wish to thank seminar and conference participants at Bonn (Max Planck Institute), Lyon (ESA), the European University Institute, Siena (Labsi) as well as James Andreoni, Uri Gneezy, Mark LeQuement, Rosemarie Nagel and Rick van der Ploeg. Karl Schlag gratefully acknowledges financial support from the Department of Economics and Business of the Universitat Pompeu Fabra, from the Universitat Pompeu Fabra, Grant AL 12207, and from the Spanish Ministerio de Educacion y Ciencia, Grant MEC-SEJ2006-09993. Joël van der Weele thanks the Nuffic grant authorities.

[†]CNRS-EconomiX, Paris and Cepremap. E-mail: galbiatir@gmail.com

[‡]Universitat Pompeu Fabra, Barcelona. E-mail: karl.schlag@upf.edu.

[§]European University Institute, Florence. E-mail: joel.vanderweele@eui.eu

Bowles (2008) for overviews). Existing explanations of this effect usually rely on the psychological concept of ‘intrinsic motivation’, a motivation to act virtuously that can be ‘crowded out’ by sanctions. In this paper we suggest another way in which sanctions can have adverse effect. It is relevant in environments where the authority with the ability to introduce sanctions is more informed than those that can be sanctioned. In particular we consider the following research questions related to the positive and negative effects of sanctions:

1. Can the incentives associated with non-deterrent sanctions induce desired behavior and make agents more *optimistic* about other players’ actions?
2. In situations of imperfect information about the past behavior of other group members, can the introduction of sanctions make agents more *pessimistic* about the actions of others by giving a signal that other players do not behave well? If so, does this reduce the effectiveness of sanctions?

We investigate these questions in an experiment based on the minimum effort game. The minimum effort game is a coordination game with many Pareto ranked equilibria. Each player chooses a level of costly effort, and is rewarded according the minimum of the efforts of all players in the group. The more efficient equilibria result only if all players play individually risky strategies. Doubt about the other player’s willingness to play such a strategy may result in inefficient outcomes. Thus, the game is particularly suitable as a workhorse to answer our questions, because there are multiple equilibria and players’ efforts are strategic complements.

Consider Question 1. Sanctions have a direct effect by providing incentives to choose higher effort. They also have an indirect or *forward looking* belief effect due to efforts being strategic complements. Anticipating that opponents are similarly affected by the sanctions and thus are expected to choose higher efforts reinforces one’s own incentive to choose a higher effort.

Question 2 addresses the signaling or *backward looking* belief effect of sanctions. When past behavior is not directly observable, sanctions may carry a signal that things are not going so well. After all, why introduce a sanction to suppress socially undesirable behavior when everybody behaves saintly? In other words, sanctions may be perceived as ‘apparently necessary’. Thus, the signaling effect of introducing sanctions may reduce the willingness to play a high and risky level of effort, and decreases the effectiveness of sanctions.

This way of looking at sanctions relates closely to an established tradition in legal scholarship. Economists and legal scholars have explored ways in which sanctions can be effective by affecting peoples’ preferences and values (Sunstein 1996, Kahan 1997, Fehr and Falk, 2002). In this role, rules and sanctions express a normative judgement of desirable behavior. Sanctions can also aid

coordination in situations where multiple equilibria exist, by changing expectations about the behavior of others (Cooter, 1998, McAdams 2000). This expressive dimension of sanctions can influence people’s behavior independently of material incentives associated with enforcement.

To answer the questions above we describe the results of a laboratory experiment, in which we focus on the effects of mild, non-deterrent sanctions in a coordination game. In particular, we look at the differences between the effects of ‘exogenous sanctions’, and the effects of ‘endogenous sanctions’ (defined below). Our workhorse game is the minimum effort coordination game with many Pareto-ranked equilibria as introduced by Goeree and Holt (2001, 2005). In all treatments agents were matched in groups of three, where the third player was a “principal” who benefitted proportionally to the minimum effort chosen by the other two in the group. The subjects played the minimum effort game twice, but the third player was the only one to be informed of the outcome of the first round before the second round was played. This information structure was common knowledge. Before the second round of the minimum effort game was played, the principal could decide whether to introduce a sanction F to both players in the group, that lowered the earnings of subject if she selected low effort. The sanction F came at a small cost to the principal’s own earnings. We call this the *endogenous* sanction, because it was introduced by a third party in a reaction to the behavior of the subjects. The sanction was ‘mild’ in the sense that it made playing low effort a more costly, but not a dominated strategy. In another treatment, the same sanction F was introduced automatically. We call this the *exogenous* sanction, because it was introduced by the experimenter unconditional on past effort choices by the subjects. Across these treatments we compare the effect of sanctions on effort choices and reported beliefs about what the other player will do.

Our results show that exogenously introduced sanctions increase beliefs about the effort that the other player will play. As a result they effectively increase coordination on more efficient equilibria. However, our answer to the second question reveals a significant difference between endogenously and exogenously introduced sanctions. In our analysis of the data we distinguish players on the basis of their behavior in the first round. The signaling hypothesis leads us to expect that people who played high effort in the first round and are confronted with a sanction, will infer that the effort of the other person must have been low. By contrast, someone who was pessimistic and played low effort will not be able to make such an inference, because she also played low, and thus a sanction may have been introduced as a reaction to her *own* behavior. We thus expect a difference between the effects of endogenous and exogenous sanctions for high effort players, but not for low effort players. In accordance with this hypothesis, we find that there is a significant difference in the effectiveness of the two kinds of sanctions for players who exerted high effort in the first round. For these players, the exogenous sanction has a substantial

positive effect on effort and beliefs about the other player's effort. By contrast, the effect of an endogenous sanctions is not distinguishable from not introducing a sanction at all. As the signaling explanation predicts, the way in which the sanction was introduced did not matter for those who played low effort in the first round.

To our knowledge this is the first paper that looks empirically at the effects of sanctions on beliefs in a minimum effort game. Moreover, it is the first paper that empirically studies the signaling effect that the introduction of sanctions may have. Its main message is that the effectiveness of sanctions depends on the context in which they are introduced. On the one hand, people recognize the incentive effects that sanctions will have on others, which multiplies their effectiveness. On the other hand, when information about the behavior of others is limited, as is the case in modern large-scale societies, the introduction of sanctions may cause pessimism by drawing attention to past misbehaviors. This is especially true for those that are optimistic and behave cooperatively. This finding implies a difficult balancing act that a government or principal must perform: It must try too keep the optimist optimistic, while at the same time encouraging the pessimists to change their behavior. The results of this experiment suggest that 'mild law' may not be the optimal way to do so, because it induces pessimism with little compensation in the way of material incentives.

A further contribution of this paper is the use of novel statistic tests. We use a new test developed by Schlag (2008) based on a so-called stochastic inequality (Cliff, 1993). This is an exact test designed to assess the direction of a treatment effect, without making (parametric) assumptions about the distribution of the samples.¹ Instead of comparing means in the underlying distribution one compares a random observation from each distribution. Note that the Wilcoxon-Mann-Whitney (WMW) test can only reject the hypothesis that two samples are drawn from identical distributions. Thus, it can identify the existence of a treatment effect, but is not informative about *why* the two distributions differ significantly. For instance, without additional assumptions, one cannot draw conclusions about whether and how the means of the samples differ. Although the results of WMW test are completely in line with our results, its different null hypothesis would have only allowed us to conclude *that* sanctions influenced behavior, we would not be able to draw conclusions how about *how* sanctions influenced behavior.

¹An exact test is a test where the statement about its level can be proven, in contrast to a level that is derived from an asymptotic approximation as the sample size tends to infinity.

2 Literature

Our experimental analysis of the effects of sanctions is related to several strands of literature. In this section we briefly summarize the links between our research and other contributions.

Our experiment relates to the literature on the crowding effects of sanctions. This term refers to the tendency of material or monetary incentives to diminish the internal motivation to engage in the desired behavior. In extreme circumstances this can lead to less of the desired behavior. This phenomenon has been empirically documented in many economic settings (see Frey and Jegen (2001) and Bowles (2008) for surveys). For our purposes, the most interesting cases involve sanctions to members of a group or a society. In a well-known experiment, Gneezy and Rustichini (2000) show that introducing a fine for picking up children late from a day-care centre resulted in an increased number of people who picked up their children late. This effect endured even after the sanctions had been withdrawn. Ostmann (1998) provides experimental results showing that external enforcement financed by experiment participants only reduces harvests in common pool problem by a small amount relative to a no-enforcement treatment. Frey and Oberholzer-Gee (1997) conducted a survey on willingness to have nuclear waste repository built in their community. Without compensation, 50.8% of the respondents answered positively, but when the request was accompanied by an offer of (substantial) monetary compensation, the acceptance rate dropped to 24.6%. Sheffrin and Triest (1992) provide evidence that tax-compliance is determined by a complex combination of attitudes and dispositions. They show that the perception that others avoid their taxes is a major determinant of the decision to evade. They also find that well publicized crackdowns on evading taxpayers may erode the trust of citizens in the other taxpayer's compliance.

Most existing explanations for the crowding effect focus on a notion of 'intrinsic motivation', which can be diminished by sanctions under certain circumstances (Frey and Jegen, 2001). Kahan (2005) suggests another explanation based on the idea that the situations in which crowding out occurs can be viewed as coordination games. Although on the face of them, settings like the ones mentioned above seem to resemble dilemma games, there is much evidence that coordination plays a large role in the outcome. This is due to the existence of so-called conditional cooperators or reciprocal agents (Fehr and Gächter, 2000). In determining their behavior in social dilemmas, conditional cooperators condition their behavior on their beliefs of what others do. Gächter (2006) surveys the evidence on the existence of conditional cooperation. Insofar as people are conditionally cooperative, their belief that others will cooperate will turn out to be a crucial variable in determining the outcome of collective action problems. Kahan argues that sanctions may provide conditional cooperators with a signal that others do not

behave well, and this diminishes their own willingness to cooperate.

There is a small theoretical literature that models this explanation of the crowding out effect. Sliwka (2007) provides a principal-agent model in which high incentives set by a principal can provide conformists in a organization with a cue that others are shirking, which may cause them to shirk themselves. Similarly, Van der Weele (2007) models a public good game where high sanctions by the government signal to conditional cooperators that defection is the social norm, which leads the government to apply lower sanctions in equilibrium.

Our research is also related to a well established strand of literature in legal scholarship: the *focal point theory of expressive law* (McAdams 2000). This view holds that laws express values and attitudes, that can shape individual behavior. Cooter (1998) argues that the expressive character of sanctions can be used to coordinate expectations on a beneficial equilibrium. People expect others to follow the law, and so a self-fulfilling equilibrium can be induced by a sanction that penalizes behavior pertaining to other equilibria. The core idea is that for this to happen, laws do not necessarily have to be fully deterrent (i.e. they can be mild), because their role is merely to create focal points. Bohnet and Cooter (2001) and McAdams and Nadler (2003) provide evidence that mild sanctions can lead to better coordination in coordination games with two equilibria. These result is in line with results about experimental coordination games showing that in coordination environments, even advisory cheap talk by an external party or coordinator can help to bring about coordination on efficient equilibria (Chaudhuri and Bangun 2007, Van Huyck et al. 1992).

Finally, we relate to the experimental literature on coordination games. The specific game that we use was introduced by Goeree and Holt (2001, 2005) who also foreshadow our answer to question 1. Unlike the present paper, they do not introduce sanctions between rounds, but investigate the behavior of different subject populations under high and low costs of effort. They show that over multiple periods, convergence to more efficient equilibria gradually takes place. Devetag and Ortmann (2007) provide a comprehensive survey of experimental results in coordination games.

In a recent paper, Brandts and Cooper (2008) compare the effectiveness of cheap talk and monetary incentives in an experimental design close to ours. Groups consist of five: four agents play a minimum effort game, and a manager profits from the degree of coordination that they reach. The manager can use financial incentives or communication messages to try to increase the level of cooperation. The authors find that communication is more effective in increasing coordination than are incentives. However, in contrast to our setup, incentives in this game cannot give any signals since the minimum effort of the previous round is known to each player.

In the paper that is perhaps closest to ours, Tyran and Feld (2006) explicitly compare the effects of endogenously and exogenously introduced mild law. In their experiment, subjects allocated to groups of three can first vote on whether sanctions for defectors should be introduced. They then play a public good game with or without the sanctions. The authors find that mild sanctions are effective when they result from the voting procedure, but not when imposed exogenously (by the experimenters). The authors show that voting for mild law raises expectations that others cooperate, and this in turn raises cooperation.

3 Discussion of the Experimental Setup

The study of sanctions comes up in settings that can often be described as either a coordination game or a Prisoners' dilemma. We chose a coordination game as an object of study, because in such games the rational choice depends only on the beliefs about the actions of the other player(s) in the game. This allows us to isolate the sanctions' effects on behavior that derive from the change in a subject's belief, and we can disregard issues to do with social preferences and/or dominant strategies that usually play a role in Prisoners' dilemmas.

3.1 The Experimental Game

We use as a workhorse the minimum effort game by Goeree and Holt (2001, 2005), because it has large action spaces that allow players to express rather precisely their preferences and beliefs. The structure of the game is as follows: two players simultaneously choose an effort level between 110 and 170 (the bounds are chosen such that there are no clear focal points). Subjects' payoffs are determined by the minimum of these two efforts, minus the cost of their own effort times a parameter $k \in [0, 1]$, which is the same for both players. In each period we also elicit from each player an interval in which he believes the other will play his effort (see below). In contrast to the original setting by Goeree and Holt (2001) in which the game is played only once, in our experiment the game is played twice. Moreover, in some treatments (see below) a sanction F was introduced in the second round, where $F = 0.5 \cdot (170 - e_i)$. Thus, F implements a subtraction to the payoffs that is proportional to the deviation of the chosen second round effort from the maximum effort (170). Although this sanction decreases the riskiness of playing higher effort, the game remains a coordination game. The sanction is applied to both the players in the group, although the actual subtraction may differ between the players depending on their second round effort choice. Another difference with the game of Goeree and Holt (2001) is the presence of a third player in the group. Depending on the treatment, this third player is either active or

inactive. When she is active, she can choose before the start of the second round whether to introduce a sanction for both players in the group. Player 3 receives a payoff proportional to the minimum effort chosen by the other two players.

In sum, payoffs in round 1 are determined as follows:

$$\begin{aligned}\pi_i(e_i, e_{-i}) &= \min\{e_i, e_{-i}\} - 0.85 \cdot e_i, \quad \text{for } i = 1, 2; \\ \pi_3(e_1, e_2) &= 0.25 \cdot \min\{e_1, e_2\}.\end{aligned}$$

where $\pi_i(e_1, e_2)$ is the payoff of player i in tokens, $e_i \in [110, 170]$ is the effort level chosen either by player 1 or player 2, and k is the cost of effort. In the second round the sanction F may be implemented by either player 3 or the experimenters. Payoffs in round 2 are given by the following equations:

$$\begin{aligned}\pi_i(e_i, e_{-i}) &= \min\{e_i, e_{-i}\} - 0.85 \cdot e_i - 0.5 \cdot (170 - e_i), \quad \text{for } i = 1, 2; \\ \pi_3(e_1, e_2, s) &= 0.25 \cdot \min\{e_1, e_2\} - s \cdot c_s,\end{aligned}$$

where c_s is the cost of introducing a sanction for the third player and $s \in \{0, 1\}$ is the choice to introduce a sanction (1) or not (0).

An important element of the experimental design is the information structure. The third player is the only one to be informed of the effort levels of players 1 and 2 when the first round is concluded. That is, at the beginning of the second round, players 1 and 2 do not know the effort levels of the other player, nor their own payoffs from the first round. However, before making any choices in the second round, players 1 and 2 know whether a sanction has been applied to their group. Note that players did not know before the first round that there would be a second round. They were informed of this only after the first round had concluded.

3.1.1 Parameters, Treatments, and Procedures

We chose to set the cost of effort at 0.85, i.e. close to 1. The evidence reported in Goeree and Holt (2001) indicates that in the presence of high costs of effort, individuals tend to coordinate on lower effort levels. We wanted effort choices to be not too high in order to give player 3 an incentive to introduce a sanction in the treatments in which she is active. We set $c_s = 4$, a level calibrated to induce roughly half of the players 3 to introduce a sanction.

We now describe the treatments. In all treatments, the first round is the same: players 1 and 2 play the minimum effort game and player 3 is inactive. In the baseline treatment there is no sanction in the second round, and player 3 is inactive. That is, the second round is conducted

exactly as the first, and no mention of a sanction was made. We refer to this treatment as the exogenous no-sanction (ExNS) treatment. By *exogenous* we mean that the choice to (not) introduce a sanction was not conditional in any way on previous decisions by the subjects. This was clear because the choice was made by the experimenters in a centralized fashion for all groups in the session. In the second treatment, sanction F is implemented in the second round. The sanction was communicated to the players before they reported their effort level and their beliefs about others' actions. They then played the second round with the sanction in place. In spirit of the experimental economic literature, we refer to the sanction in neutral terms, i.e. as a "subtraction". In the remainder, we refer to this treatment as the exogenous sanction treatment (ExS).

Although player 3 is present in all treatments, she is only active in the third treatment. After player 3 has observed the chosen effort levels of players 1 and 2 in the first round, she is asked to decide whether to a) change both player 1's and 2's payoff structure in the second round by introducing a sanction F , or b) leave the payoff structure unaltered with respect to the first round. After player 3 has taken her decision, players' 1 and 2 are informed of it. They then play the second round with payoff structure decided by the principal. We refer to this treatment as either the endogenous sanction treatment or EnS (if a sanction is introduced by player 3) or the endogenous no-sanction treatment or EnNS (if no sanction was introduced).

Because the experiment features just two rounds of play and no possibility of learning, it was very important that people understood the game correctly from the start. To this purpose we ran a tutorial before the start of the first round. In the tutorial, participants had 5 minutes to come up with hypothetical effort choices of players 1 and 2 and to calculate their payoffs resulting from these choices. The tutorial took place before assigning subjects to a role, so that also players 3 could practice with the calculation of payoffs of players 1 and 2. In addition to this tutorial, the input screens in the actual experiment provided subjects with the possibility calculate their payoffs from a given choice. That is, after entering and before confirming their choices, subjects could enter a hypothetical choice of the other player and let the computer calculate their payoffs resulting from these choices.

The experiment was conducted in several sessions at the economics lab of the university of Siena, Italy. The first sessions took place in May and June 2007. Another series of sessions was conducted in November 2007. Subjects entered their effort and belief choices on a computer that was running on the software z-Tree (Fischbacher, 2007). The number of subjects in an experimental session varied between 18 and 30. The subjects earnings were in tokens as specified above, which were converted into euro's at the end of the experiment at an exchange rate of 10 tokens = 0.75 euro. The instructions were read out loud to make them public knowledge. The

instructions and the input screen are provided in appendix B.

3.1.2 Elicitation of a Belief Interval

Apart from the effort choices, we are interested in the effect of sanctions on players' anticipation of what the other will do. Therefore, in the same input screen in which players 1 and 2 enter their effort choice, we asked them to enter beliefs about the other player's effort choice in that round. Rather than elicit a point belief, we decided to elicit an *interval*. More precisely, players have to specify a range (i.e. a lower bound L and its upper bound U) in which the other player's choice is believed to fall. In order to increase accuracy in belief reporting we reward a correct guess². The earnings from a guess are determined as follows:

$$\pi_i(L, U) = \begin{cases} 0 & \text{if } e_{-i} \notin [L, U] \\ 0, 15 \cdot (60 - (U - L)) & \text{if } e_{-i} \in [L, U] \end{cases}$$

That is, a wrong guess (the actual number chosen by the other player falls outside the specified range) yields no payoff. A correct guess (the actual number chosen by the other player lies within the specified range) yields 15% of difference between the length of the interval [110, 170] and the width of the interval $[L, U]$. Thus, the smaller the specified range, the higher the earnings if the guess is correct. However, a smaller range also increases the risk that the guess is not correct, in which case no tokens are earned.

Eliciting an interval has the advantage that it gives information not only about the location of the belief distribution, but also about its dispersion. Schlag and Van der Weele (2009) show that provided this is single peaked, this *interval scoring rule* will induce rational decision makers to include both the median and the mode of their belief distribution in the chosen interval. Moreover, the width of the interval determines an upper bound on the variance of the belief distribution. This makes the width of the interval a proxy for how 'sure' the decision maker is. These results hold for any risk neutral or risk averse decision maker. Note that the alternative quadratic scoring rule is only guaranteed to reveal the mean when the decision maker is risk neutral.

4 Non-Parametric Tests of Stochastic Inequality

One contribution of this paper is the use of new non-parametric tests that have been designed for small samples (Schlag, 2008). The disadvantage of existing tests is that they either add

²Gächter and Renner (2006) show that incentivizing beliefs' reporting has a positive impact on beliefs accuracy.

distributional assumptions (e.g. assuming normality or restricting the parameter space so that the alternative hypothesis is no longer the complement of the null hypothesis) or that they can only establish that a treatment changes the distribution of outcomes, not how. Specifically, the standard non-parametric test in the experimental literature for comparing samples has been the Wilcoxon-Mann-Whitney (WMW) test. The null hypothesis of this test is that the two samples are drawn from the same population. Thus, unless one is willing to make further assumptions on the underlying distributions (i.e. that all other moments of the probability distributions except the mean are equal), the WMW test cannot identify the direction of the treatment effects. It can only establish that they are different.

We analyze the effect of sanctions by testing ‘stochastic inequality’. In order to identify the direction of a treatment effect we compare the likelihood that one variable realizes a higher outcome than the other. We measure this degree by the so-called *stochastic difference* which ranges from -1 to 1 . Specifically, given two random variables Y_1 and Y_2 , $\delta = Pr(Y_2 > Y_1) - Pr(Y_2 < Y_1)$ is called the stochastic difference of Y_1 versus Y_2 . δ is estimated by taking the sample average across all pairings of the data. One says that Y_2 tends to realize higher outcomes than Y_1 if $\delta(Y_1, Y_2) > 0$. To establish a treatment effect in this direction, we test the null hypothesis that $\delta \leq 0$. When $Pr(Y_1 = Y_2) = 0$ then this is equivalent to testing that $Pr(Y_2 > Y_1) \leq 1/2$. When the data is given as matched pairs then the appropriate test is a sign test. When data is given by two independent samples, we implement the test of Schlag (2008). Appendix A gives a more extensive formal treatment of these procedures.

It is worth noting that there are no other exact nonparametric tests for comparing means or testing stochastic inequality given independent samples. In particular, the WMW test is not an exact test for comparing the underlying means given two independent samples (e.g. see simulations of Forsythe *et al.*, 1994). Neither are there other exact nonparametric tests for correlation; the Spearman rank correlation test can only identify non-identical distributions. Non-exact tests of stochastic inequality have appeared in the biostatistical applications (Brunner and Munzel, 2000). One innovation of the tests we use here is that they are exact, in the sense of having the level that they are claimed to have, and do not rely on asymptotic approximations. They are the first exact tests for this stochastic inequality based on independent samples. Unlike tests for means, the ordinal nature of tests of stochastic inequality makes them less sensitive to outliers and hence they are very well suited to uncover significant differences given small samples.

We want to emphasize that the results of the more traditional Wilcoxon-Mann-Whitney test support our analyses. All significant results that we present are also significant, often more so, in the corresponding WMW test (these results are available on request). However, as

explained above, without further assumptions the null hypothesis of the WMW test does not allow us to draw conclusions about the direction of the effect. Because the WMW test is rather powerful, we will use it when we want to gather support for a claim that two samples have similar distributions. In this case we are not primarily interested in the direction of the effect. Rather, we want to have the strongest possible test to falsify the claim that two samples are similar.

5 Hypotheses and Results

In this section we present the results of our experiment. We present our analysis by testing conjectures that are based on the research questions mentioned in the introduction. These conjectures are specific enough to provide us with the null hypotheses necessary for classical statistical analysis.

5.1 Statistics for the Entire Sample

The number of participants in the experiment was 243: 45 in treatment 1, 51 in treatment 2, and 147 in treatment 3. In treatment 3, the principal decided to introduce a sanction in 29 out of 49 groups. Each experimental session lasted roughly 35 minutes and the subjects earned 7.5 euros on average³. In the tutorial 82% (199 out of 243) correctly computed the payoffs from hypothetical choices. As another indication of whether people understood the game, we also checked whether there were ‘anomalous observations’: people who specified an effort choice above the upper bound of their belief interval. We found just 6 such observations.

We observe a high correlation between beliefs and effort in the first round of each treatment, as you would expect in a minimum effort game. The correlation coefficient between the lower bound of beliefs and the effort choice is 0.85^{***}, which is highly significant⁴. The correlation with the upper bound was somewhat lower (0.81^{***}), because many subjects specified an upper bound at, or close to 170 in the first round. They were thus restricted in moving this upper bound in the second round. For most subjects this was not true for the lower bound of the belief interval. For this reason we take the lower bound of the interval as our indicator of beliefs throughout this paper.

Figure 1 shows a histogram of first period effort choices, aggregated over all treatments. We

³If this seems little, remember that the incentives were concentrated on only two (effort) choices. At each of these choices there was thus relatively a lot at stake.

⁴The significance is based on a test with the null hypothesis that the covariance is less than 0 (Schlag, 2008)

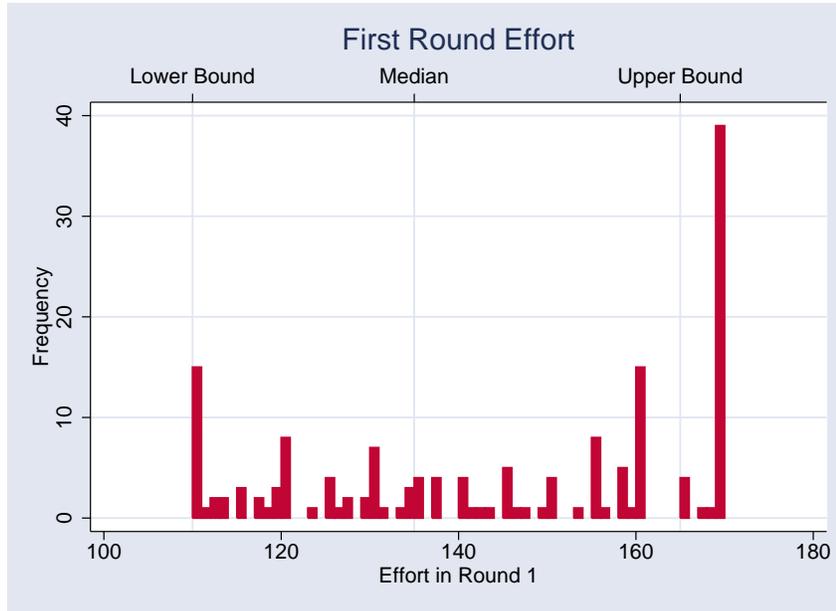


Figure 1: Histogram of first round effort choices of all subjects.

see a large clustering of observations around 170, a smaller cluster around 110 and an otherwise fairly uniform distribution⁵. We want to analyze the effect of the introduction of a sanction in the second round, and hence in the remainder we focus on the *changes* of effort and beliefs between rounds. We compute for each subject the change in beliefs and effort levels, and compare these changes across treatments.

There are a few complications to analyzing changes between rounds. First of all, the observations for the group members in the third treatment are not independent. The effort decision of one subject in the first round will influence the decision to implement a sanction by the third player. This in turn may influence the effort and beliefs of the other subject in the second round. When we do statistical testing, we correct for this dependence by taking the average of two observations whenever the subjects come from the same group, and treating it as one observation.

Second, interpreting changes in efforts and belief intervals as reaction to the experimental setting is not straightforward. Subjects that specified an effort level or a lower bound on beliefs of (or close to) 170 in the first round are unable to adjust this level upwards, and subjects who chose close to 110 cannot adjust it further downwards. This will generate observations of zero

⁵The effort levels are higher than those in Goeree and Holt (2001) with a cost of effort of 0.9. Reasons may be that the cost of effort is slightly lower in our setup and that in the instructions we did not use the word “cost” when referring to k .

changes that may not reflect the actual preferences or adjustment of beliefs of the participants.

As we will see, the general trend in the experiment was for subjects to adjust their beliefs and efforts upwards in the second round. Thus, the problem is not severe for those who are initially on the lower bound. Specifically, there were no subjects who chose low effort (below 135 but above 110) and subsequently moved their effort downwards, and only three who chose a (small) downward adjustment of beliefs. Therefore we do not consider those who chose 110 to be severely constrained. However, the matter is different for those who chose effort or belief levels on (or close to) the upper bound of 170. It is likely that most of those subjects would have liked to change their behavior if they had been able to move upward further, but were constrained to do so. We believe that the fact that these people do not change their behavior does not give us accurate information about their actual change in beliefs and their preferences over effort levels. Therefore, we restrict ourselves to analyzing the choices of those subjects who actually had a choice. We focus on comparing the behavior across treatments of subjects who reported beliefs or effort lower than or equal to 165. In practice this means that for the analysis of the beliefs, we excluded subjects who chose first round belief levels strictly higher than an upper bound of 165. This resulted in excluding 11 observations. For the analysis of the efforts, we excluded subjects who chose first round effort levels strictly higher than 165. This resulted in the exclusion of 39 observations. The median first round effort of the sample thus obtained is 135. The values of the upper and lower bound that we applied are indicated in 1. In the remainder, we define high effort players as those who play first round effort in $e \in \{135, \dots, 165\}$ (i.e. above the median), and low effort players as those who play first round effort in $e \in \{110, \dots, 134\}$ (i.e. below the median).

5.2 Effort and Beliefs in the Baseline Treatment (ExNS)

While our analysis will focus on comparing behavior across treatments it is of interest to consider what happens in the baseline case, where there are no exogenous sanctions. Recall that there is no feedback between rounds in the treatment without sanctions. One might conjecture that in the absence of feedback there is no change in effort and yet it is not clear whether behavior should not change over time simply due to the fact that a choice is made a second time. We present the evidence in Table 1. We denote by Mean ExNS1 the mean of first round variables in the exogenous no-sanction treatment, and by ExNS2 the second round variables. The last column presents the estimated stochastic difference of the first round versus the second round in treatment ExNS (remember from section 4 that this is the estimate of $\delta = Pr(Y_2 > Y_1) - Pr(Y_2 < Y_1)$). It is worthwhile to note that testing for stochastic inequality for matched pairs is equivalent

	n	Mean ExNS1	Mean ExNS2	Stochastic Difference ExNS1 vs ExNS2
Effort	23	133	137	0.17
Belief	29	134	138	0.15**

Table 1: Mean efforts, mean beliefs, and stochastic difference between round 1 (ExNS1) and 2 (ExNS2) in the exogenous no-sanction treatment (ExNS). * Denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

to performing a sign test. We find insignificant differences in the effort (confirmed by the Wilcoxon rank sum test). On the other hand we find significant evidence that the lower belief level tends to be higher in the second round. Apparently people move up their belief levels, but as we can see from Table 1, changes are small so people are not sufficiently optimistic to change their effort levels by much.

5.3 The Effect of Exogenous Sanctions (Question 1)

Our first question relates to the effects of exogenous sanctions on efforts. In the case of exogenous sanctions we can abstract from any signaling considerations because the sanction is unconditionally imposed by the experimenters. Sanctions are modeled in our experiment by an additional cost of making efforts below the maximum 170. Mathematically this translates into a reduced cost of effort. Under a given belief distribution such a change in the cost of effort causes a rational agent to increase effort.

If the subject anticipates that the other player also increases effort, her beliefs about opponent effort become more optimistic, which makes it rational to increase effort even more. Thus, we expect that introducing sanctions causes an increase in beliefs but an even stronger increase in effort. If we compare behavior in round one and round two in the sanction treatment, we cannot separate this anticipated effect of sanctions from other effects that we observed in the case of no sanctions. The appropriate benchmark for comparison is the treatment without sanctions. We formulate the following conjecture about this comparison:

Conjecture 1 *The change in effort and belief levels between rounds 1 and 2 is larger when there are exogenous sanctions than when there are no exogenous sanctions in period 2. This effect is more pronounced for efforts than it is for belief levels.*

This conjecture can also be motivated with the results of Goeree and Holt (2001), who find

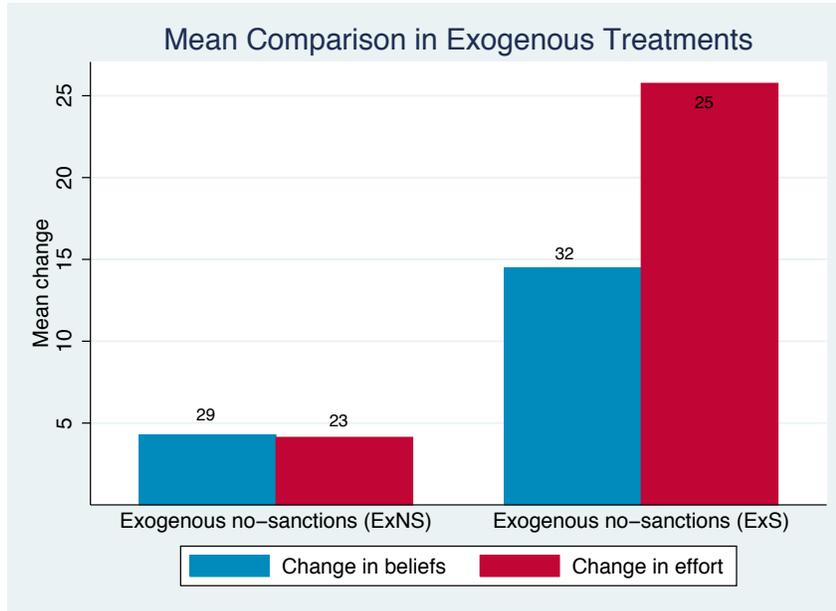


Figure 2: The change in beliefs and sanctions for the whole sample, except those who chose first round efforts $\in \{166, 167, \dots, 170\}$ or first round beliefs $\in \{166, 167, \dots, 170\}$. (Number of independent observations for each sample at the top of the bar).

that a lower cost of effort increases effort levels in a between-subject design. Since our sanction effectively lowers the cost of effort, it is reasonable to conjecture that (exogenous) sanctions will increase beliefs and effort. We now gather evidence for our conjecture. Figure 2 presents the change in the means between round 1 and round 2 for the ExS and ExNS treatments. In Table 2 we report the results of our statistical analysis of conjecture 1. We estimate the stochastic difference of the change in effort under exogenous sanctions (ExS) versus the change in effort under exogenous no-sanction (ExNS). To indicate changes between the two rounds of a treatment X we use the notation dX . Similarly we consider the changes in the lower bound of the belief intervals, comparing the change under exogenous sanctions, and the change under no exogenous sanction. Comparing the significance levels we indeed observe a more pronounced difference in terms of effort than in terms of beliefs, as is also apparent from Figure 2. To formally test this finding would involve designing a new test which is outside the scope of this paper. However we do note that the 20% equi-tailed confidence intervals overlap; by this crude method at least this difference is not found to be significant.

Summary 1 *We confirm our conjecture that changes in efforts and beliefs tend to be higher when there are exogenously imposed sanctions in the second round than when there are no sanc-*

	Stochastic Difference dExNS vs. dExS
Effort	0.64***
Belief	0.31*

Table 2: Values of stochastic difference between *changes* in the exogenous no-sanction (ExNS) treatment and *changes* in the exogenous sanction (ExS) treatment. * Denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

tions in the second round. The data lend support to the claim this effect is stronger for effort than for beliefs.

5.4 The signaling Effect of Sanctions (Question 2)

We now investigate the effects of endogenous sanctions. We compare subjects' choices under exogenous sanctions to subjects' choices under endogenous sanctions. Note however that there are at least two differences between these two groups. One difference is that in the exogenous case the sanction was imposed by the experimenter while in the other case it was imposed by a subject in the experiment. A second difference arises from the fact that the choice of a sanction by the subject need not be unconditional (like the experimenter's sanction) or random. The choice of a sanction may reflect the observations of particular first round effort choices. It is exactly this kind of information transmission we wish to analyze, and the experiment is designed to isolate the signaling effect from the incentive effects of sanctions, by comparing ExS and EnS. Before we analyze the reactions of the subjects to the imposition of an endogenous sanction, we investigate the choice of sanction by the third player.

5.4.1 The Choice of Endogenous Sanctions

To see why player 3 would decide to implement a sanction, consider her monetary incentives. The third player is rewarded proportionally to the minimum group effort. However, imposing sanctions carries a small cost. A maximizing principal will implement a sanction if she expects to recoup these costs through an increased minimum effort level. When initial effort is low, there is a large potential range for effort increases, and changing behaviors can be very profitable. Moreover, if effort is low in the first round, there is no clear reason to think that it will rise without a sanction. Thus we can formulate the following conjecture:

Conjecture 2 *In the endogenous sanction treatment, the likelihood of sanctions being imposed by the ‘principal’ is decreasing in the minimal effort chosen in the first round.*

In order to test this conjecture we compare the minimum first round effort in the sanctioned groups to the minimum first round effort of non-sanctioned groups. We use the Wilcoxon-Mann-Whitney test because we are interested in any difference between the samples. However, we cannot find marginally significant evidence that the distributions of minimal effort are different in the groups where sanctions are imposed as compared to the group without sanctions imposed (the p -value is 0.63). Of course the samples are small, so the test is not very powerful. However, as Table 5.4.1 shows the descriptive data do not point at large differences either. Note that

	Mean of Min. Group Effort	# Below 165	# Above 166
No Sanction	138	17	3
Sanction	135	28	1

Table 3: Descriptive data on first round minimum effort of sanctioned and non-sanctioned groups. The columns show the mean, and the number of groups with minimum effort below 165 and above 166.

sanctions were also introduced occasionally when minimum effort was high. Note that this need not contradict equilibrium behavior. To see this, assume that there are some subjects that always choose low effort (‘low’ types) while others choose high effort as they believe that the others that think like them also choose high effort. There can be equilibria in which a sanction is imposed only if minimum effort is high, and therefore are a signal that the group consists of high types. Observing no sanction be a signal that the other subject is of type low and hence it would be best to choose low. Thus the principal will impose sanctions on high types to preserve coordination. This behavior is optimal for all players, provided there are sufficiently few low types to make play of high effort in the first round an equilibrium. Obviously there are other equilibria in which coordination on high effort is not sanctioned. This multiplicity may be a reason why there is no clear pattern when sanctions are imposed. For all practical purposes however, we can just assume that the behavior is random. This leads us to the following conclusion:

Summary 2 *We have no significant evidence that sanctioned groups had lower minimum effort. The descriptive statistics similarly indicate a lack of a clear pattern. Sanctions seem to be randomly imposed in our data set.*

This result implies that there is no endogeneity problem that could have arisen if only low-effort players had been sanctioned. To the extent that people who play low effort react different to sanctions than others, this would have made the comparison with exogenous sanction treatment more difficult. To this comparison we turn now.

5.4.2 The Effect of Endogenous Sanctions

Although the apparently random imposition of sanctions means that there is no clear informational content of sanctions, subjects may still believe that sanctions were imposed systematically. Specifically, subjects may follow the same reasoning that led us to formulate Conjecture 2. If this is the case, sanctions may still influence beliefs about the other group member. A small thought exercise teaches us that the inference that can be made depends on a subjects' own effort in the first round. Consider a subject who believes Conjecture 2 to be true. Assume first that this subject chose high effort in the first round. When she observes that the principal imposes no sanction, the subject infers that the opponent chose a high effort because otherwise they would have been sanctioned. This may give her cause for optimism, and a reason to keep choosing high effort. On the other hand, if the high-effort subject is sanctioned, she infers that it is likely that the opponent made a low effort. The high effort player will face the following questions: Will the opponent react to the sanction with a sufficient increase in effort such that I should increase my own effort too? Or is the opponent simply someone with a tendency to make low efforts even under sanctions, in which case I should lower my own effort? Compared to the case of exogenous sanction, the observation of a sanction induces uncertainty that the other subject chose low effort and will do so again. Now assume that the subject played low effort in the first round. A sanction no longer has any informational content as long as the subject believes in Conjecture 2. Specifically, any sanction can always be interpreted as being aimed at the subject himself. Thus, there is no reason to assume his beliefs about the opponent will change, and we expect him to behave much like someone under exogenous sanctions would behave.

Note that higher order expectations that the players may have about each other may complicate this pattern. For example, the low-effort player who observes a sanction may think that if his opponent is a high-effort player, she will now be discouraged. We content ourselves with trying to identify first-order patterns. We summarize these patterns in two conjectures (remember from Section 5.1 that by low first round effort we mean effort $\in \{110, \dots, 134\}$, and by high effort we mean effort $\in \{135, \dots, 165\}$).

Conjecture 3 *a) For those that chose a low effort in the first round, the change in efforts*

and beliefs under endogenous sanctions will be similar to the change under exogenous sanctions.

- b) For those that chose a high effort in the first round, the change in efforts and beliefs will be larger under exogenous sanctions than under endogenous sanctions (signaling effect).

We first consider Conjecture 3a). Figure 3 presents the mean changes in beliefs and effort for people who played low effort in the first round. Figure 3 reveals no large differences between the

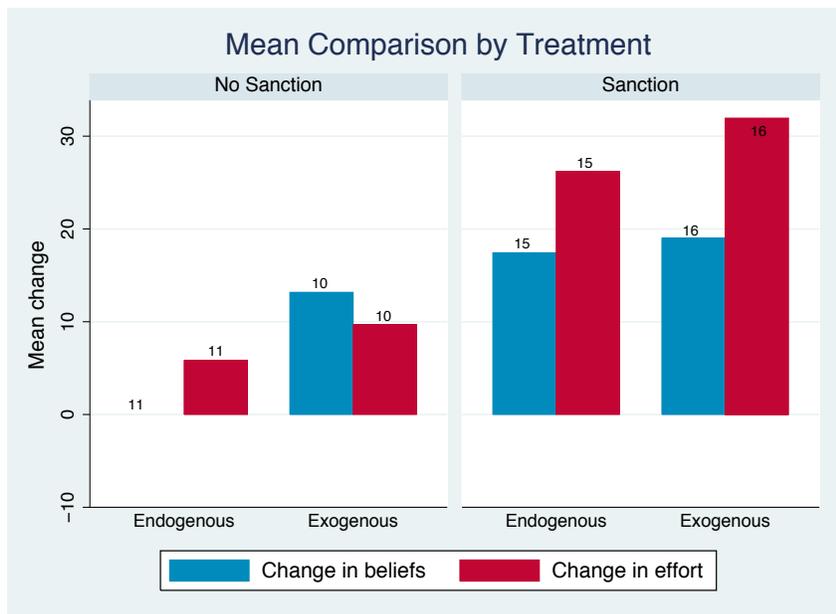


Figure 3: Means of changes in beliefs and effort across treatments, for those who played low effort ($\in \{110, 111, \dots, 134\}$) in the first round. (Number of independent observations for each sample at the top of the bar).

exogenous and endogenous sanction treatments. We now try to falsify Conjecture 3a). We test the null hypothesis that the distribution of change in effort is identical in the endogenous and exogenous sanction settings. Since we are interested in any difference between the distributions we use the WMW test. The results in Table 4 show that we cannot reject the null hypothesis of identical distributions in the exogenous and endogenous treatments, both for effort and beliefs.

The problem is that the sample sizes are small, so we can only provide limited evidence of similarity.⁶ Therefore, we will now show that we can make similar claims about the effectiveness of sanctions, regardless of the way they were introduced. We compare first and second round

⁶Using statistical hypothesis testing we can show at most that the differences are not too large, since formally it is impossible to obtain significant evidence that the effect of endogenous and exogenous sanctions is equal.

	WMW p -values dEnS vs. dExS
Effort	0.29
Belief	0.97

Table 4: p -values of the Wilcoxon Mann-Whitney rank sumtest of the exogenous and endogenous treatments for those who played low effort ($\in \{110, 111, \dots, 134\}$) in the first round. * Denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

efforts and beliefs between the exogenous and the endogenous treatments, both for sanction and no sanction. We report results in Table 5.

	Stochastic Difference	
	ExS1 vs. ExS2	EnS1 vs. EnS2
Effort	1***	1***
Belief	0.5**	0.8***

Table 5: Estimates of stochastic difference between round 1 and round 2 of treatments ExS and EnS, for those who played low effort ($\in \{110, 111, \dots, 134\}$). * Denotes significance at 10%, ** denotes significance at 5%, *** Denotes significance at 1%.

We find very similar estimates of stochastic difference in both sanction treatments. We feel confident therefore to draw the following conclusion:

Summary 3 *For subjects that made low efforts in the first round we find no significant evidence that endogenous and exogenous sanctions have different effects on either efforts or beliefs.*

We will now test conjecture 3b). In Figure 4 we report average changes in efforts and beliefs across treatments for subjects who played high efforts ($\in [135, 165]$) in the first round. Eyeballing the figure, it seems like the exogenous sanctions are more effective than the endogenous ones for those who played high effort. The results based on stochastic differences, reported in Table 6, confirm this. We observe significant evidence that exogenous sanctions are more effective in raising effort than endogenous sanctions. There is marginal significant evidence that beliefs tend to change more under exogenous sanctions. One wonders whether endogenous sanctions have any effect at all. To find out we test if there is a difference between the endogenous sanction

However, the larger the sample size, the more powerful the test, and the more confident we are that the effect, if it exists, is small.

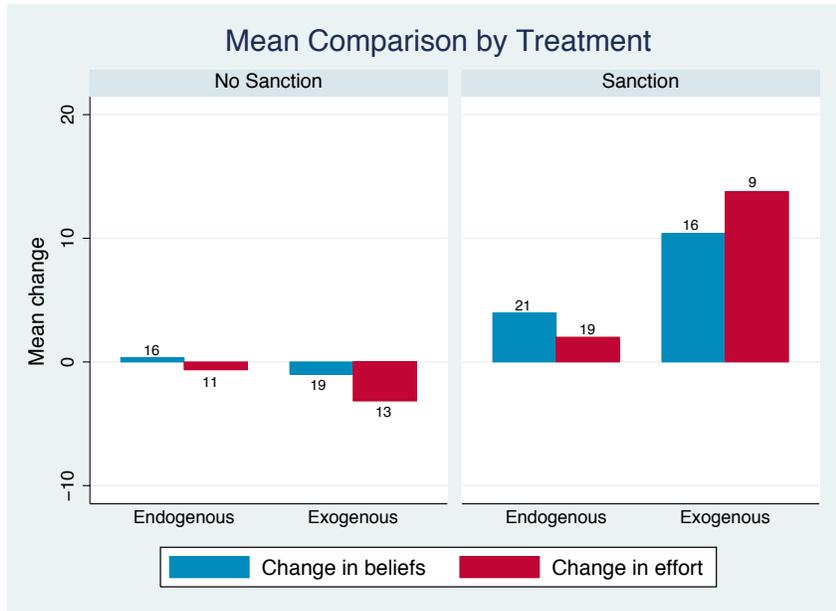


Figure 4: Means of changes in beliefs and effort across treatments, for those who played high effort ($\in \{135, 136, \dots, 165\}$) in the first round. (Number of independent observations for each sample at the top of the bar).

	Stochastic Difference dEnS vs. dExS
Effort	0.66**
Beliefs	0.39*

Table 6: Estimates of stochastic difference between the exogenous and endogenous sanction treatments for those who played high effort ($\in \{135, 136, \dots, 165\}$) in the first round. * Denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

treatment and the baseline treatment (ExNS). In the first column of Table 7 we report the p -values of the WMW test for this comparison.

	WMW p -value dExNS vs. dEnS	Stochastic Difference dExNS vs. dExS
Effort	0.35	0.78***
Belief	0.49	0.48**

Table 7: Comparison of the baseline (ExNS) treatment and the sanction treatments for those who played high effort ($\in \{135, 136, \dots, 165\}$) in the first round. * Denotes significance at 10%, ** denotes significance at 5%, *** denotes significance at 1%.

We find that endogeneity dampens the increase in efforts and beliefs. In fact, it dampens it so much that the effect of endogenous sanctions cannot be distinguished from not mentioning and introducing sanctions at all. However, the sample sizes are small, so it is possible that we would not be able to reject the null hypothesis of equal distributions, even if the actual difference is quite large. To counter this criticism, the second column of Table 7 shows the comparison with the baseline treatment with the *exogenous* sanction. It is clear that for similar sample sizes we get very significant results of the effectiveness of exogenous sanction.

Summary 4 *For subjects who played high effort in the first round, endogenous sanctions are less effective in raising efforts and beliefs than exogenous sanctions. In fact, the effect of endogenous sanctions cannot be distinguished from the effect of not introducing a sanction at all.*

5.5 Belief Intervals

Before we move to the conclusions, we investigate the results pertaining to the width of the belief interval $U - L$. One of the reasons we asked the participants to specify an interval rather than a point belief was that we are interested in the impact of sanctions on uncertainty about the behavior of the other player, for which the size of the interval $U - L$ is a proxy (see Schlag and van der Weele, 2009). Figure 5 shows the changes in the width of the belief interval for those who chose the lower belief interval in $\{110, 111, \dots, 165\}$ in the first round). As Figure 5 shows, uncertainty did not change between rounds in both no-sanction treatments, while uncertainty went down in both sanction treatments. We can confirm this result with statistical analysis. Table 8 presents the estimates of stochastic difference between the first and the second round interval width in all treatments. In both no sanctions cases a test of stochastic inequality cannot reject the null hypothesis that the distributions in the two rounds are equal at the

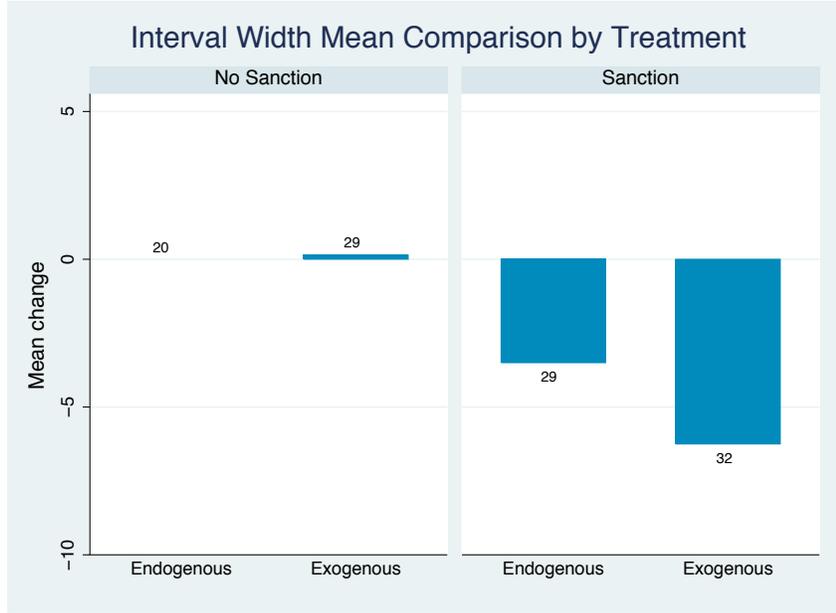


Figure 5: Means of change in the width of the interval across treatments, for those who chose the lower belief interval in the first round in ($\in \{110, 111, \dots, 165\}$) in the first round (number of independent observations for each sample at the top of the bar).

10% level. By contrast, we find that there is significant evidence that the interval decreases under exogenous sanctions and marginally significant evidence that the interval decreases under endogenous sanctions. This reinforces our conclusion that sanctions facilitate coordination partly by reducing uncertainty about the behavior of others.

If sanctions were to have a signaling effect, we would expect for those subjects who chose high effort ($\in \{135, 136, \dots, 165\}$) in the first round, that the reduction in uncertainty is smaller under endogenous sanctions than under exogenous sanctions. Testing the direction of the effect with stochastic inequality, we find that the tendency of the decrease in uncertainty is in fact significant at 1% in the exogenous sanction treatment, while under endogenous sanctions it is no

	Stochastic Difference			
	EnNS1 vs. EnNS2	ExNS1 vs. ExNS1	EnS2 vs. EnS2	ExS1 vs. ExS2
Interval Width	0.067	0.0	-0.31*	-0.46**

Table 8: Estimates of stochastic difference between the round 1 and round 2, for those who chose the lower belief interval in the first round in ($\in \{110, 111, \dots, 165\}$) in the first round. * Denotes significance at 10%, ** denotes significance at 5%, *** Denotes significance at 1%..

longer significant. Moreover, the estimates for stochastic inequality do not reveal a significant difference between EnS and either ExNS or EnS. This indicates that endogenous sanctions do not reduce uncertainty for those who played high effort in the first round relative to the no-sanction treatments. It thus seems that sanctions reduce uncertainty in general, except for endogenous sanction applied to those who played high effort. This is congruent with our signaling explanation. However, when we directly compare the change in the interval width between both sanction treatments for those who played high effort, we cannot reject the null hypothesis of no difference.

Summary 5 *Uncertainty about the choice of the other player, as measured by the width of the belief interval, declines in the sanction treatments. There is no evidence of a change in the no-sanction treatments. For those who played high effort in the first round, the reduction in uncertainty only occurs under exogenous sanctions.*

6 Discussion and Conclusion

The results of our experiment allow us to conclusively answer our two questions. Over the whole sample, exogenous sanctions clearly have a positive effect on effort levels and beliefs about others' effort level. However, the way in which sanctions are introduced matters. This manifests itself in the fact that for people who played relatively high effort in the first round, the difference between the effect of an endogenously and an exogenously introduced sanction is significant. In fact, the endogenously introduced sanctions cannot be distinguished from the treatment without (exogenous) sanctions.

We think that the most plausible rationale for this result is the idea that underlies our hypotheses. The endogenous introduction of sanctions gives subjects a signal that the other group member did not 'cooperate', in the sense that she selected low effort. This tends to make people more pessimistic about the effort played by their companion in the group and less willing to move up in effort themselves. For those who played high effort initially this pessimism is reflected in the fact that beliefs and effort do not significantly increase under endogenous sanctions. We also found that uncertainty, as measured by the width of the belief interval, does not go down under endogenous sanctions as it does under exogenous sanctions. A signaling effect also explains why the difference between the sanction treatments does not occur for people that play low effort in the first round. For them this signaling effect is less pronounced, because they may think that the sanction was aimed at them rather than at the other player in the group.

Our results discredit a naive view of deterrence in which it is only the economic incentives that

matter for behavior. The literature on crowding and intrinsic motivation had already established that sanctions may have adverse effects in some situations. We have identified another reason why sanctions may be ineffective. The result supplies a motivation why ‘mild law’ may not work. In contrast to Tyran and Feld (2006), we provide evidence that the *endogenous* introduction of sanctions rather than the exogenous one may be the cause of problems. In Tyran and Feld, a voting procedure for the introduction of a mild sanction gives people the opportunity to send a public signal that they are willing to cooperate. This in turn leads to increased cooperation. In our experiment, the introduction is under the discretion of a third player who has observed past play of the game. This setup reflects more closely the arrangements of a society where people make the laws through representatives, rather than directly. In this case a sanction sends exactly the opposite signal: sanctions are apparently necessary to keep people from deviating from the efficient outcome. The results show not only that such an effect can exist when the information conditions are right, but also that it is potentially quite substantial. Our study thus suggests that mild law may not be the best instrument in this case, because it does not compensate for this signaling effect by providing adequate incentives for efficient behavior.

In our experiment we observed the fact that the signaling effect was not present for low effort players, because the groups were so small that the sanction was likely to reflect their own behavior. However, in real life, relevant communities consist of many more than two people. This means that even people who play low effort may interpret the sanction as a signal, because it is unlikely that a sanction is introduced on the basis of the behavior of one person. Assuming some external validity of the experiment, one can conclude that a sanctioning authority needs to attain a careful balance between correcting the behavior of deviants or pessimists and maintaining the optimistic beliefs of cooperators. The results of the paper have implications for both public policies and manager-employees relationships in firms. As pointed out by Brandts and Cooper (2006), coordination failure can cause corporations and other organizations to become trapped in unsatisfactory situations both for managers and employee.

How to attain such a balance is an interesting further research question that goes beyond the aim of this paper. One possibility is to try to avoid the issue altogether by implementing harsh laws making undesired action very costly. Such a deterrent law would presumably override the signaling effect. However, such laws and their enforcement may be costly to implement in the real world, since they require at least some probability of detection for undesired activity and potentially costly sanctioning activities. Another possibility to investigate is whether appropriate framing of the introduction of a law can mitigate the signaling effect. In the tradition of experimental economics, this paper has tried to use neutral framing, replacing “effort” with “a number”, and “sanction” with “subtraction”. In real life however, a policy maker could attempt

to surround the introduction of sanctions by soothing or stimulating messages. For example, one may say the actual number of people who deviate from the efficient strategy is small, or express the expectation that they will conform to the sanction. However, it is theoretically unclear why such cheap talk would be effective. The experiments by Brandts and Cooper (2008) and Van Huyck et al. (1992) incorporate the possibility of a principal to send written messages and suggestions to the agents. These studies could be combined with the asymmetric information structure in this paper in order to study this issue.

Last but not least, we wish to push forward the use of exact tests that “let the data speak” and do not add distributional assumptions. One approach in the experimental literature on crowding out has been to use the Wilcoxon-Mann-Whitney test to uncover differences in distributions, and to complement this test by looking at the descriptive statistics to make statements about the direction of the effect. A more popular approach throughout the experimental literature has been to implicitly use the WMW test as test for comparing means, without mentioning the condition needed for its validity, namely that all moments of the distributions except the first have to be the same. A contribution of our paper is the use of new tests (that are exact but do not impose additional distributional assumptions) that allow us to test directly for a negative impact of sanctions. We think these tests are an important addition to the toolbox of economists working with small data sets.

7 References

- Bohnet, Iris, and Robert Cooter (2001). “Expressive Law: Framing of Equilibrium Selection?” Mimeo, Harvard University and UC Berkley.
- Bowles, Sam (2008). “Policies Designed for Self-Interested Citizens May Undermine the Moral Sentiments: Evidence from Economic Experiments”, *Science*, 320.
- Brandts, Jordi and Cooper, David (2008). “It’s What You Say Not What You Pay”, forthcoming in *Journal of the European Economic Association*.
- Brandts, Jordi and Cooper, David (2006). “A Change Would Do You Good: An Experimental Study on How to Overcome Coordination Failure in Organizations”, *American Economic Review*, 96(3), 669-693.
- Brunner, Edgar and Munzel, Ullrich (2000). “The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation”, *Biometrical Journal*, 42, 17-25.

- Chaudhuri, Ananish, & Bangun, Laura (2007). “Credible Assignments in the Minimum Effort Coordination Game”, Mimeo.
- Cliff, Normann (1993). “Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions”, *Psychological Bulletin* 114, 494-509.
- Cooter, Robert (1998). “Expressive Law and Economics”, *Journal of Legal Studies*, 27, 585-607.
- Devetag, Giovanna and Ortmann, Andreas (2007). “When and Why? A Critical Survey on Coordination Failure in the Laboratory”, *Experimental Economics*, 10(3), 331-344.
- Fehr, Ernst and Falk, Armin. (2002). “Psychological Foundations of Incentives”, *European Economic Review*, 46, 687-724.
- Fehr, Ernst and Simon Gächter. (2000). “Fairness and Retaliation: The Economics of Reciprocity”, *Journal of Economic Perspectives*, 14(3), 159-81.
- Fischbacher, Urs (2007). “z-Tree: Zurich Toolbox for Ready-made Economic experiments”, *Experimental Economics* 10(2), 171-178.
- Forsythe, Robert, Horowitz, Joel L., Savin, N. E., and Sefton, Martin (1994). “Fairness in Simple Bargaining Experiments”, *Games Economic Behavior*, 6, 347-369.
- Frey, Bruno S. and Jegen, Reto (2001). “Motivation Crowding Theory”, *Journal of Economic Surveys*, 15(5), 589-611.
- Frey, Bruno S., & Oberholzer-Gee, Felix (1997). “The Cost of Price Incentives: an Empirical Analysis of Motivation Crowding-Out” *American Economic Review*, 87(3), 746-755.
- Gächter, Simon (2006). “Conditional cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications”, CeDEx Discussion Paper, 2006-03.
- Gächter, Simon and Elke Renner (2006). “The Effects of (Incentivized) Belief Elicitation in Public Good Experiments”, University of Nottingham.
- Gneezy, Uri and Rustichini, Aldo (2000). “A Fine is a Price”, *Journal of Legal Studies*, 29, 1-17.
- Goeree, Jacob and Holt, Charles A. (2005). “An Experimental Study of Costly Coordination”, *Games and Economic Behavior*, 51, 349-364
- Goeree, Jacob and Holt, Charles A. (2001). “Then Little Treasures of Game Theory and Ten Intuitive Contradictions”, *American Economic Review*, 91(5), 1402-22.

- Kahan, Dan M. (2005). "The Logic of Reciprocity: Trust, Collective Action, and Law," in *Moral Sentiments and Material Interests*. Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr eds. Cambridge, Massachusetts: MIT Press.
- McAdams, Richard H. (2000). "A Focal Point Theory of Expressive Law", *Virginia Law Review*, 86, 1849-1731.
- McAdams, Richard H. and Nadler, Janice (2003). "Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game", *Journal of Empirical Legal Studies*, 2(1), 87-123.
- Ostman, Andreas, (1998). "External Control May Destroy the Commons", *Rationality and Society* 10 (1), 103-122.
- Schlag, Karl H. (2008). "A New Method for Constructing Exact Tests without Making any Assumptions", Working Paper 1109, Department of Economics and Business, Universitat Pompeu Fabra.
- Schlag, Karl H. and Joël J. van der Weele, (2009) "An interval scoring rule", Mimeo, European University Institute. Available from <http://www.eui.eu/Personal/Researchers/joelvdweele/Work/Work.html>.
- Sheffrin, Steven M. and Triest, Robert K. (1992). "Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance", in *Why People Pay Taxes*, J. Slemrod ed.
- Sunstein, Cass R. (1996). "On the Expressive Function of Law", *University of Pennsylvania Law Review* 144(5), 2021-2053.
- Tyran, Jean-Robert and Feld, Lars P. (2006). "Achieving Compliance when Legal Sanctions are Non-deterrent," *Scandinavian Journal of Economics*, 108(1), 135-156.
- Van Huyck, John, Gillette R. and Battalio, Raymond (1992) "Credible Assignments in Coordination Games", *Games and Economic Behavior*, 4, 606-626.

Appendix A: Stochastic Difference and Inequality

Given two random variables Y_1 and Y_2 , $\delta(Y_1, Y_2) = \Pr(Y_2 > Y_1) - \Pr(Y_2 < Y_1)$ is called the *stochastic difference* of Y_1 versus Y_2 . The stochastic difference can be estimated by computing the sample analogues. Consider first the case of matched pairs where data is given by joint observations of Y_1 and Y_2 . The estimate is calculated by ignoring all pairs in which $Y_1 = Y_2$ and then taking the difference between the empirical frequency of pairs with $Y_2 > Y_1$ and of pairs in which $Y_2 < Y_1$. Now consider the case in which there are two independent samples, one associated to each variable. Here one can estimate δ by considering the frequency of $Y_2 > Y_1$ among all possible pairs and subtracting from this the frequency in which $Y_2 < Y_1$ among all these pairs. The resulting estimates are unbiased.

If $\delta(Y_1, Y_2) > 0$ then one says that Y_2 tends to yield larger outcomes than Y_1 . We wish to identify significant evidence that Y_2 tends to yield larger outcomes than Y_1 . So we wish to test the null hypothesis $H_0 : \delta(Y_1, Y_2) \leq 0$ against the alternative hypothesis $H_1 : \delta(Y_1, Y_2) > 0$ for a given specified level α . This is called a test of *stochastic inequality* (Cliff, 1993, Brunner and Munzel, 2000).

Assume that data has the form of *matched pairs* as given by n independent observations of (Y_1, Y_2) . Then this test reduces to a sign test. One uses a binomial test to test whether the probability that $Y_2 > Y_1$ conditional on $Y_2 \neq Y_1$ is $\leq 1/2$.

Now assume instead that data is given by two independent samples of Y_1 and of Y_2 . Let n_i be the number of observations of Y_i , $i = 1, 2$. We present an exact test of these hypotheses due to Schlag (2008).

Randomly match one observation of each sample to generate $\min\{n_1, n_2\}$ matched pairs. Then determine a rejection probability based on the randomized version of the sign test with size $0.2 \cdot \alpha$. The combination of the matching and the probabilistic recommendation yields an exact randomized test with size $0.2 \cdot \alpha$. We proceed as follows to derive an exact nonrandomized test that has level α . Reject the null hypothesis if the rejection probability of the above randomized test is above 0.2. Note that the factor used to reduce the size of the randomized test is equal to the threshold used to translate the randomized recommendation into a deterministic recommendation.

Appendix B: Instructions

We report instructions for the endogenous sanction treatment.

Originally in Italian

Instructions for the first round

Introduction

Welcome! You are going to take part in an experimental study of decision making. Please follow these instructions carefully. You will be paid according to your performance. At the end of the experiment we will tell you how much you earned.

Once everyone is seated we will formally start the experiment by reading the instructions. After this reading you will have the opportunity to ask us questions about the procedure. However at no time may you communicate with any of the other participants of your session. Please also refrain from talking to others about your experience until tomorrow in order not to influence others taking part in our experiment. Please turn off your mobiles in case they are still switched on. We hope you have fun.

Matching and assignment to a role

The computer will assign you by chance (i.e. at random) to a group consisting of three participants. You will not know the identity of the other two in your group and they will not know your identity. The computer will also assign a role to each in this group. Two of this group (from now on: player 1 and player 2) will have to take a decision as described below, the third (from now on: player 3) will be inactive but still will earn some money.

Decisions and Earnings

During the experiment any choice will lead to some earnings expressed in tokens. Total earnings at the end of the experiment are determined by the sum of all earnings and will then be converted into money at the exchange rate of

1 token = 7.5 Eurocents (or equivalently: 100 tokens=7,5 Euro)

It will not be possible to have negative earnings at the end.

Player 1 and Player 2

Players 1 and 2 will simultaneously each be asked to make two decisions: to choose a number and to make a guess about which number the other player chooses. Both decisions have to be entered into a decision screen that is described in more detail below. Neither player will observe the decisions of the other player.

Choosing a Number

Both player 1 and player 2 have to choose a number. This number can be any number between and including 110 and 170 (fractions or decimals not allowed).

The earnings in tokens of either player 1 or player 2 from choosing a number are determined as follows. A player receives the lower of the two numbers chosen by player 1 and player 2 minus 85% of their own number.

This has the following implications:

- Assume players 1 and 2 chose the same number. Then a player will receive his/her own number (since both numbers are equal, this is also the lowest number) minus 85% of his/her own number.

- Assume that players 1 and 2 chose different numbers. Then, the player who chose the lower number could have increased his earnings by choosing a slightly higher number. However, the player who chose the higher number could have increased his earnings by choosing a slightly lower number.

The following mathematical representation will not be read out loud.

Suppose (among players 1 and 2) that one of them chooses the number Y and the other chooses the number Z .

If $Y = Z$ then the player who chose Y receives $Y - 0.85 \times Y$.

If $Y < Z$ then the player who chose Y receives $Y - 0.85 \times Y$.

If $Y > Z$ then player who chose Y receives $Z - 0.85 \times Y$.

In addition, players 1 and 2 first receive a fixed amount of 35 tokens.

Guessing the other's choice

In addition to specifying a number, both player 1 and player 2 are asked to make a guess about the number chosen by the other player. The guess is made by specifying a range (given by its lower bound L and its upper bound U) in which the other player's choice is believed to belong.

The earnings in tokens of either player 1 or player 2 from making this guess are determined as follows. A wrong guess (the actual number chosen by the other player falls outside the specified range) yields nothing. A correct guess (the actual number chosen by the other player lies within the specified range) yields 15% of the difference between 60 and the width of the range $U-L$. Therefore the smaller the specified range, the higher the earnings if the guess is correct. However, a smaller range also increases the risk that the guess is not correct, in which case no tokens are earned.

The following mathematical representation will not be read out loud:

If the number Z chosen by the other player lies in the range (it is greater than or equal to L

and less than or equal to U) then the player who has chosen L and U gets $0.15 \times (60 - (U - L))$ tokens if this number Z does not lie within the range then the player who has chosen L and U gets nothing.

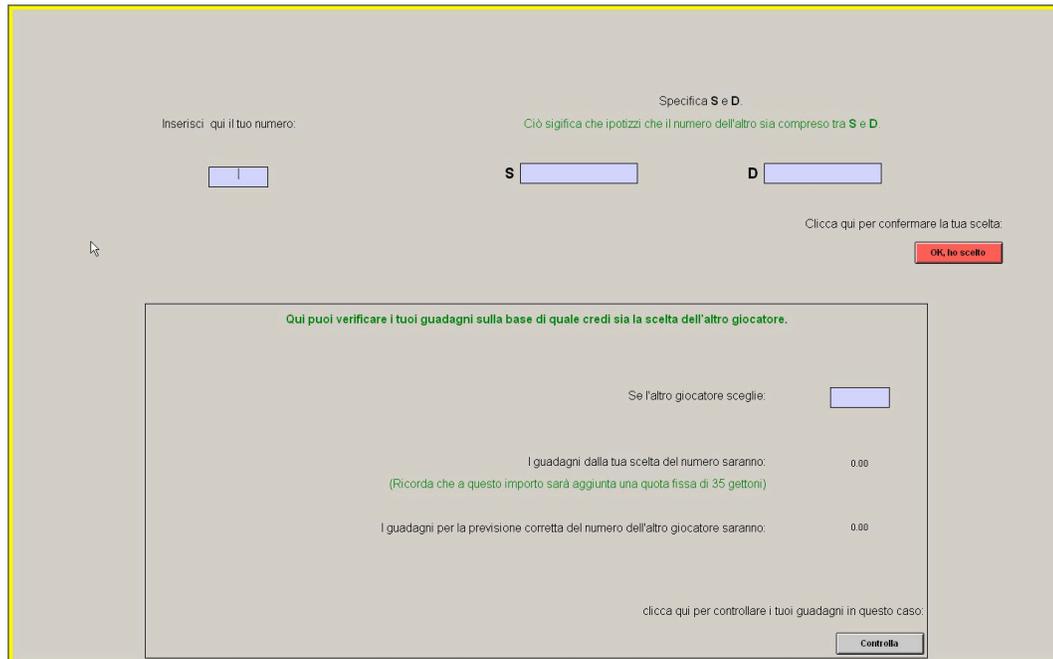


Figure 6: Input screen in the first round.

Player 3

Player 3 does not make any decision during the experiment and earns an amount of tokens equal to 25% of the smaller of the two numbers chosen by players 1 and 2.

A more mathematical representation of this statement will not be read out loud:

Tokens earned by player three = $0.25 \times$ (smaller of the two numbers chosen by player 1 and player 2)

Tutorial

Before the experiment starts, so before roles are assigned, all participants have the possibility to practice and to get used to the structure of the game. To this end, you will participate in a tutorial round, where you will see the decision screen as described above. You will have 5 minutes to enter as many different values as you like for both your own number and your guess,

and the other player's hypothetical number. You can then use the check button to see what your earnings from these numbers and your guess would be. You are encouraged to verify the calculation behind the earnings of both the number choice and the guess. The values entered in this tutorial have no influence on your earnings and will not be recorded. After 5 minutes the tutorial will stop and the experiment will start.

Final Remarks

During the experiment, you are not permitted to speak or communicate with the other participants. If you have a question while the experiment is going on, please raise your hand and one of the experimenters will come and answer it.

At this time, do you have any questions about the instructions or procedures? If you have a question, please raise your hands and one of the experimenters will come to your seat to answer it.

Instructions for the second round

Introduction

Now we run a second and final experiment. Earnings will be added to your previous earnings. After this new experiment everything is over and your total payment will be calculated.

This new experiment is very similar to the previous one up to some changes we highlight.

Matching and roles

All participants are matched with the same people as before and keep the roles they had before.

Decisions and Earnings

IN CONTRAST to the previous experiment, player 3 now also makes a decision.

Player 3

At the start of the experiment, before player 1 and 2 make any decisions, player 3 observes the numbers chosen by players 1 and 2 in the previous experiment. After having observed these numbers, player 3 makes a decision that determines how earnings of players 1 and 2 are calculated in this new experiment. The outcome of this decision is observed by players 1 and 2 before they make their choices. Player 3 has the following two choices:

a) NOT CHANGE: To choose "not change" means that the earnings of all players are as in the previous experiment. In particular, player 3 earns 25% of the smaller of the two numbers chosen by players 1 and 2.

b) **CHANGE:** To choose “change” means that earnings in tokens of all players are changed as follows. Players 1 and 2 receive the lower of the two numbers chosen minus 85% of their own number minus 50% of the difference between 170 and the player’s own chosen number. . That is, relative to the previous experiment, there is an extra amount subtracted to your earnings that is larger the smaller your number is. Player 3 earns 25% of the smaller of the two numbers chosen by players 1 and 2 minus 4. The terms that are new as compared to the previous experiment have been underlined.

Mathematical illustration not to be read out loud:

Suppose player 3 chooses “change” and (among players 1 and 2) that one of them chooses the number Y and the other chooses the number Z .

If $Y = Z$ then the player who chose Y receives $Y - 0.85 \times Y - 0.5 \times (170 - Y)$.

and player 3 receives $0.25 \times Y - 4$.

If $Y < Z$ then the player who chose Y receives $Y - 0.85 \times Y - 0.5 \times (170 - Y)$.

and player 3 receives $0.25 \times Y - 4$.

If $Y > Z$ then player who chose Y gets $Z - 0.85 \times Y - 0.5 \times (170 - Y)$

and player 3 receives $0.25 \times Z - 4$.

Regardless of the choice of player 3, player 1 and 2 also receive a fixed amount of 35 tokens.

Player 1 and Player 2

As in the previous experiment, players 1 and 2 make two decisions: choose a number and make a guess by specifying a range. Earnings from making the guess are as in the previous experiment, earnings from choosing a number are specified above.

Input Screens here

Final Remarks

If you have any questions then please ask them now.

Please do not log off the computer when the experiment is over.