Journal of Economic Behavior & Organization 94 (2013) 34–51

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Sanctions that signal: An experiment☆

Roberto Galbiati[a,*], Karl H. Schlag[b], Joël J. van der Weele[c]

[a] OSC-CNRS and Sciences-Po, Paris, France
[b] University of Vienna, Austria
[c] J.W. Goethe University, Grüneburgplatz 1, RuW Gebäude 4, Stock, 60323 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form 27 May 2013
Accepted 1 August 2013
Available online xxx

JEL classification:
C92
D83
K42

Keywords:
Sanctions
Beliefs
Expressive law
Deterrence
Coordination
Minimum-effort game

ABSTRACT

The introduction of sanctions provides incentives for more pro-social behavior, but may also be a signal that non-cooperation is prevalent. In an experimental minimum-effort coordination game we investigate the effects of the information contained in the choice to sanction. We compare the effect of sanctions that are introduced exogenously by the experimenter to that of sanctions which have been actively chosen by a subject who has superior information about the previous effort of the other players. We find that cooperative subjects perceive actively chosen sanctions as a negative signal which significantly reduces the effect of sanctions.

1. Introduction

Authorities commonly take measures in order to promote cooperation between people, including laws, sanctions, and monitoring devices. These interventions provide incentives for good behavior, but the very fact that they are introduced can also change people's perception of the organization or society they are a part of. For example, increasing punishment for a particular crime can inform the public that this crime is prevalent and hard to control. Introducing time management systems in a workplace may signal that shirking is the social norm. Increasing monitoring on immigrant groups may lead people to believe that these groups have bad intentions and have a stigmatizing effect. International financial intervention in a country can inform investors that its mismanagement has been worse than previously thought.

In all these examples, the introduction of the intervention sends a signal that others are not cooperating, which may dampen or even reverse the desired impact of the measure. Recently, a number of theorists have modeled such a signaling effect, based on particular assumptions on how beliefs are formed (Sliwka, 2007; Friebel and Schnedler, 2011; Bénabou and

☆ For valuable comments and discussions we wish to thank numerous seminar and conference participants as well as James Andreoni, Uri Gneezy, Mark Le Quement, Rosemarie Nagel, Emeric Henry, Steffen Huck 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.

* Corresponding author. Tel.: +33 (0) 1 45 49 56 36.
E-mail addresses: galbiatir@gmail.com (R. Galbiati), karl.schlag@univie.ac.at (K.H. Schlag), vanderweele@econ.uni-frankfurt.com (J.J. van der Weele).

0167-2681/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jebo.2013.08.002

Tirole, 2011; Van der Weele, 2012). Our objective is to understand how people actually make inferences in this context, and to this end we run a laboratory experiment. More specifically, we study behavior in an experimental coordination game, designed to answer the following questions.

1 Can the introduction of incentives associated with small, non-deterrent[1] sanctions induce efficient behavior and raise expectations of cooperative actions by other players?
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 implicitly signaling that other players have not been cooperating? If so, does this reduce the effectiveness of sanctions?

Our experimental setup is a two person minimum-effort game: a coordination game with many Pareto ranked equilibria, based on the setup in Goeree and Holt (2001, 2005). Each player chooses a level of costly effort, and is rewarded according to 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.

In all experimental 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 principal 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. Apart from effort choices, we also elicited the subjects' beliefs about the effort of the other player.

To answer Question 1, we compare a control treatment without sanctions with a sanction treatment. In a control treatment no sanctions were introduced between rounds, and consequently the second round was the same as the first. In the treatment, a mild sanction F was introduced for both players in the group, that lowered the earnings of a subject if she selected low effort (and also carried a small fixed cost for the principal), but did not change the set of Nash equilibria. Because the sanction was introduced by the experimenter unconditional on past effort choices by the subjects, we call this the 'exogenous sanction treatment'.

Our hypotheses for the effect of such sanctions are based on a simple formal model, explained in Section 4. The experimental game has many Nash equilibria, so in order to make predictions we use a model similar to level-k reasoning (Nagel, 1995; Costa-Gomes and Crawford, 2006). We assume that people think of their partner as either a pessimistic type, who believes the partner will choose low effort, or an optimistic type, who believes their partner will choose high effort. This model predicts that the change in payoffs associated with sanctions should increase effort through (a) an 'incentive effect', and (b) a 'belief effect', i.e. a change in expectations about the action of the other player. The first result of the paper is that we do indeed find both effects in the data, and so our answer to Question 1 is affirmative.

Question 2 addresses a potential signaling effect of sanctions. To answer it, we introduced an additional treatment. Before the second round of the minimum-effort game was played, the principal could decide whether or not to introduce the same sanction F as above, at a small cost to his own earnings. Because the principal had observed first round behavior and could condition the sanction on this behavior, we call this the 'endogenous sanction treatment'.

Our model predicts that in this treatment there exists an equilibrium in which the principal will sanction if and only if there is at least one player who plays relatively low effort. Therefore, a player who played relatively high effort in the first round, but nevertheless observes a sanction, will learn that his partner is a pessimistic type who is likely to continue to choose low effort: sanctions are 'bad news'. This means that for these players the incentive effect of the sanction will be counteracted by a negative belief effect. By contrast, people who initially chose low effort will not get any information from a sanction, because in equilibrium it would have been introduced independently of their partner's behavior. Thus, our hypothesis is that endogenous sanctions are less effective than exogenous sanctions, especially for those who behaved cooperatively.

Our second result is that we can confirm these hypotheses. There is no evidence that subjects with low first-round effort react differently when facing an endogenous sanction. On the other hand, the signaling effect of the endogenous sanction for those with high first-round effort is so strong that it eliminates the incentive effect, so the net effect is indistinguishable from the case where there is no sanction. As a result, exogenous sanctions are on aggregate significantly more effective than endogenous sanctions.

To our knowledge, we are the first to investigate experimentally whether the introduction of sanctions signals uncooperative behavior by other group members. The main message of our paper 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 increases their effectiveness. On the other hand, when information about the behavior of others is limited, as is the case in modern large-scale societies or firms, 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.

[1] By 'small', 'non-deterrent' or 'mild', we mean that sanctions do not make playing the socially efficient action a dominant strategy.

Finally, this paper makes two methodological contributions. First, we use novel tests that can correctly identify significant evidence that sanctions increases effort without making distributional assumptions. Second, we use a new, incentive compatible mechanism to elicit belief intervals.

2. Literature

Our experimental study relates to several strands of the literature. First, our paper is inspired by a theoretical literature in economics on the potential signaling that occurs when imposing a sanction, or more generally when introducing some policy. Bénabou and Tirole (2003) show how the choice of incentives can provide information to an agent about the difficulty of a task. This paper has recently been tested experimentally by Bremzen et al. (2011). Sliwka (2007), Van der Weele (2012), Bénabou and Tirole (2011) and Friebel and Schnedler (2011) investigate how the choice of incentive policies can signal information of the policy maker about the relative prevalence of different types in the population to imperfectly informed agents. In each of these papers, the signaling effect of sanctions depends on the existence of agents with different preferences. In equilibrium, sanctions are a signal that there are many selfish types around, which reduces the motivation of the agents to exert effort, either because of conformist preferences or because of complementaries in technology. The common finding in this literature is that the signaling effect of sanctions leads the principal to use sanctions less often relative to situation where the agents are perfectly informed. We complement this theoretical literature by showing that the signaling effect can also obtain in a setting in which all agents have identical preferences, and only differ according to their beliefs.

Second, we contribute to the growing experimental literature on 'endogenous sanctions'. A recent literature has contrasted the effect of sanctions for defection in a public good game that are introduced exogenously by the experimenter, with similar sanctions that are implemented by direct voting mechanisms. These studies show that sanction schemes introduced through voting can raise expectations that other group members will be cooperative. As a consequence, and in contrast to our findings, endogenous sanctions are more effective than exogenous sanctions (Tyran and Feld, 2006; Markussen et al., 2011; Kamei, 2010).

Our paper is also related to field experiments on the crowding out effect of sanctions (see Frey and Jegen, 2001; Bowles, 2008 for a survey). For example, Frey and Oberholzer-Gee (1997) show empirically that a monetary incentive lowers acceptance rates of nuclear repository waste. Gneezy and Rustichini (2000) show that a fine for picking up children late from a day-care center actually increased late-coming. Although the signaling mechanism we describe may be at work, these papers cannot differentiate it from potential alternative explanations, such as a direct impact of incentives on preferences, or the idea that a fine is a signal about some relevant characteristic of the principal. By contrast, our explicit distinction between exogenous and endogenous sanctions and the use of a between-subject design allows us to identify that sanctions carry signals about the past behavior of other agents.

Thirdly, we contribute to the experimental literature on the effect of incentives in coordination games (see Devetag and Ortmann, 2007 for a survey of experimental results in coordination games). Goeree and Holt (2001, 2005) find in a between-subject design that effort levels in a minimum-effort game are higher when effort costs are lower. By contrast, our Question 1 refers to the effect of the introduction of sanctions, hence refers to a within subject design. Brandts and Cooper (2006) look at the effect of exogenous bonuses that are proportional to the minimum group effort, whereas the size of the sanction we consider is based on individual choices and therefore depends on the individual's behavior only.

Finally, Xiao (2013) also studies a signaling effect of sanctions by superiorly informed third parties. In a sender-receiver game, an outside 'enforcer' can punish a sender who sends false messages to the receiver. Because the payoffs of the enforcer do not depend on whether the sender deceives, it is perhaps not so surprising that sanctions become a signal of the sender's deception. By contrast, in our study, the interests of the players and the enforcer are aligned. Thus sanctions have a dual role of both signaling information and enhancing coordination. It is the tension between these two roles that is the focus of this paper.

3. Experimental setup

The theoretical literature on the signaling effect of sanctions described above assumes that there is heterogeneity in the preferences of agents. Moreover, information about the preferences of others is relevant for behavior due to the existence of technological complementarities or social interdependencies (e.g. conformism). In this paper, we choose a slightly different approach by selecting a simple coordination game with multiple equilibria as an object of study. This allows us to investigate directly the effect of sanctions on inferences about the other player's behavior, as opposed to the preferences underlying that behavior. We can also minimize the effects of social preferences since all players have the same ranking over equilibria.[2]

[2] Social preferences such as an altruistic concern for other player's payoffs could still play a role, but their influence is likely to be less pronounced than in a prisoner's dilemma or public good game, where there is an obvious conflict of interest between the players.

3.1. The experimental game

As the coordination game underlying our experiment we choose the minimum effort game of Goeree and Holt (2001, 2005). Large action spaces help capture variation in players' beliefs. In this game, two players simultaneously choose an action, to be interpreted as an effort level, between 110 and 170 (the bounds are chosen such that there are no clear focal points). Subjects' payoffs are equal to the minimum of these two efforts, minus the amount of their own effort times a cost parameter $k \in [0, 1]$, which is the same for both players.

While in the original setting by Goeree and Holt (2001) the game is played only once, in our experiment the game is played twice where treatments differ according to what happens in the second round. In some treatments a value F is subtracted from the payoffs in the second round, where $F = 0.5 \cdot (170 - e_i)$. The subtraction of F can be interpreted as a sanction, since deviations from the maximal effort (170) are punished proportionally. The sanction is 'mild', as the game remains a coordination game with the same set of pure strategy Nash equilibria.

Another difference to Goeree and Holt (2001) is that we include a third player who is either active or inactive, depending on the treatment. When active, the third player can choose before the start of the second round whether or not to introduce a sanction for both players in the group. When inactive, the third player does not make any choice, instead the choice of whether to introduce a sanction is made by the experimenter. Regardless of her activity status, player 3 receives a payoff proportional to the minimum-effort chosen by the other two players. Note that player 3 is present in each treatment to maintain the same context.

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, $e_i \in [110, 170]$ is the effort level chosen by player i, $i = 1, 2$, and $k = 0.85$ is the cost of effort. Payoffs in round 2 are given by the following equations:

$$\begin{aligned} \pi_i(e_i, e_{-i}, s) &= \min\{e_i, e_{-i}\} - 0.85 \cdot e_i - s \cdot 0.5 \cdot (170 - e_i) \quad \text{for } i = 1, 2, \\ \pi_3(e_1, e_2, s) &= 0.25 \cdot \min\{e_1, e_2\} - 4s, \end{aligned}$$

where 4 is the cost of introducing a sanction for the third player and $s \in \{0, 1\}$ reflects whether a sanction was introduced ($s = 1$) or not ($s = 0$).

An important element of the experimental design is the information structure. Players do not know before the first round that there will be a second round. They are informed of this only after the first round has concluded. Furthermore, players 1 and 2 do not observe each other's effort levels, nor do they learn their own payoffs before both rounds are over. When active, player 3 is informed about the effort levels of players 1 and 2 in round 1. Players 1 and 2 only observe before round 2 starts whether or not player 3 has chosen to introduce a sanction.

3.1.1. Parameters, treatments, and procedures

The experiment was conducted at the economics lab of the University of Siena, Italy between May and November 2007, using the software z-Tree (Fischbacher, 2007). Within each session subjects were matched into groups of 3, faced the same treatment and played the two round game described above a single time. Before playing the game the instructions were read out loud and a tutorial was conducted. The subjects received a show up fee of 1 euro, their 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.

We set the cost of effort $k = 0.85$, i.e. rather high to induce low effort choices. Sanctioning should be moderately costly for player 3 so that there is sufficient diversity in the choices of player 3. Accordingly we chose to set the cost for the third player of introducing a sanction equal to 4, which is comparable to a reduction of $4/0.25 = 16$ in the minimal effort of players 1 and 2.

We now describe the treatments. Instructions can be found at in Appendix E. Treatments are all the same in the first round: players 1 and 2 play the minimum effort game and player 3 is inactive. In the control 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. In particular subjects are not aware of the fact that sanctions are introduced in other treatments. We refer to this treatment as the 'exogenous no-sanction' (ExNS) treatment. In the treatment we refer to as the 'exogenous sanction' (ExS) treatment, the sanction is introduced in the second round. Players 1 and 2 are told that the term F is subtracted from their payoffs; player 3 remains inactive, as in ExNS. We use the term *exogenous* to indicate that introduction of a sanction is not conditional in any way on previous decisions by the subjects. This was clear to the subjects because all subjects that belonged to the same session received the same instructions and this was common knowledge as instructions were read publicly.

The 'endogenous' treatment is the only one in which player 3 has an active role. After round 1, player 3 observes the effort levels chosen by players 1 and 2 in the first round, whereupon she is asked to decide whether to (a) introduce the sanction F to the payoffs of players 1 and 2 at a cost 4 to her own payoffs, or (b) to leave the payoff structure unaltered. After player 3 has taken her decision, players 1 and 2 are informed of it and choose their effort levels. With some abuse of

the word 'treatment' we refer to the case where player 3 did (not) introduce the sanction as the endogenous (no) sanction treatment (EnS, EnNS).

Because the experiment features just two rounds of play, it was very important that people understood the game correctly from the start. For this purpose, before subjects were assigned to a role, we ran a tutorial where participants had 5 min to choose hypothetical effort choices of players 1 and 2 and to calculate their payoffs resulting from these choices. In addition to this tutorial, the input screens in the actual experiment provided subjects with the possibility to calculate their payoffs from a given set of choices.

3.1.2. Elicitation of a belief interval

Because we are interested in the subjects' beliefs, we ask players 1 and 2 about their beliefs about the effort of their partner in each round. In a minimal effort game, it is not just the expected effort level of the other player that is of interest, but also the downward deviations from this level. Moreover, we are interested in how uncertainty differs between treatments and rounds. We capture both of these features of beliefs by eliciting confidence intervals about the partner's effort.

More precisely, players 1 and 2 have to specify an interval (i.e. a lower bound L and an upper bound U) in which the effort chosen by the other player is believed to fall. Elicitation is remunerated as follows:

$$\pi_i(L, U, e_{-i}) = \begin{cases} 0 & \text{if } e_{-i} \notin [L, U] \\ 0.15 \cdot (60 - (U - L)) & \text{if } e_{-i} \in [L, U] \end{cases}$$

where $\pi_i(L, U, e_{-i})$ is the payoff of player i who specifies a range $[L, U]$ when e_{-i} is the effort chosen by the player matched with player i. Note that a smaller interval increases the payoff of a correct guess but also increases the risk of not being correct and obtaining no tokens. Schlag and van der Weele (2011) show that this rule incentivizes subjects who are risk neutral or risk averse to specify an interval that contains the partner's effort choice with a probability of at least 50%. Moreover, increased uncertainty about the partner's effort (a more dispersed belief distribution) leads to the specification of a wider interval.

In the following, we focus on the lower bound of the belief interval since variation is largest for this variable, but results are similar if we use the midpoint of the interval instead. Thus, in the following, the term belief refers to the value of L. In addition, to understand how uncertainty changes between rounds and differs between treatments we consider the width of the interval $U - L$.

4. Hypotheses

The hypotheses we present in this section are based on a simple model. We believe that the intuitions from this model are relatively straightforward, so we relegate a formal treatment of the model to Appendix A.

In contrast to the theoretical literature on this topic, which assumes heterogenous preferences of agents, we base our hypotheses on a model where there is heterogeneity in beliefs only. Since the game has many equilibria, we use a level-k thinking model (Nagel, 1995; Costa-Gomes and Crawford, 2006) to generate predictions. Specifically, we assume that each player assumes that the other player (or 'partner') is best responding to one of two distinct belief distributions that are fixed over time. As a consequence, players believe that their partner chooses one of two effort levels which we refer to as 'high' effort and 'low' effort.

Consider first choices in round one. Following the fact that subjects did not know that there would be a second round, we ignore strategic considerations with respect to round two. For simplicity, players are assumed to be risk neutral. Therefore, in round one, a player will choose between the same two effort levels that the partner is believed to be choosing. In particular, the high effort is chosen if and only if the probability that the partner is choosing this effort is sufficiently high.

Consider now round two. As choices will depend on the treatment, we first look at exogenous sanctions. The sanction reduces the marginal cost of effort which results in both an incentive and a belief effect. The incentive effect describes the change in behavior that results if beliefs would remain unchanged. Lower marginal effort costs means that some players have an incentive to switch from low to high effort while none will switch from high to low effort. The belief effect refers to the change in behavior driven by players changing their beliefs about their partner's behavior. Players anticipate that their potential partner will exert higher effort (through the incentive effect) which additionally increases their own incentive to increase their effort. Because each player assumes that the partner best-responds to fixed belief distributions, no further iterations in strategic reasoning are necessary.

It follows that an exogenous sanction will increase both own effort and the beliefs about the effort of the other player. Under a mild assumption on the consistency of beliefs in round one (as specified in the appendix), the average effort of the players increases more than their average beliefs do. The reason for this is that effort is increased through both the incentive and the belief effect, whereas beliefs are only affected by the latter. This leads to our first hypothesis.

Hypothesis 1. Exogenous sanctions increase beliefs and efforts. This effect is more pronounced for effort levels than it is for belief levels.

Consider now the treatment with endogenous sanctions. We wish to determine when the third player, or principal, will choose to sanction and how players 1 and 2 will react to the (non-)introduction of sanctions. We focus on behavior that can be sustained in a Perfect Bayesian Nash Equilibrium for any prior beliefs of the principal about the beliefs of the agents. We

find three candidate equilibrium behaviors for the third player: "always sanction", "never sanction" and "sanction when at least one player exerts low effort". For instance, unconditional sanctioning is best if sanctions raise effort sufficiently to offset the cost for the third player. For this to happen, the third player has to anticipate an increase in minimum effort by at least 4/0.25 = 16 points as a result of the sanction. The equilibrium involving "sanction when at least one player exerts low effort" exists when (i) a player with low effort is sufficiently responsive to a sanction and (ii) the difference in effort between low and high effort players is sufficiently large. The intuition behind these equilibrium conditions will become clearer below.

Common to these equilibrium strategies is that low efforts are always sanctioned if sanctions are chosen. This leads to our next hypothesis.

Hypothesis 2. In the endogenous treatment, the likelihood of sanctions being imposed by the principal is decreasing in the minimal effort chosen in the first round.

We now turn to the reaction of players one and two to the sanctioning choice, which naturally depends on the principal's equilibrium policy. In the "always sanction" or "never sanction" equilibrium, no information about partner behavior is transmitted to the players. In this case, endogenous sanctions have the same effect as exogenous sanctions. Information about partner behavior is transmitted only under "sanction when at least one player exerts low effort", and it is only transmitted to a player who exerted high effort. In this equilibrium, absence of a sanction reveals to a player with high effort that the matched partner exerted high effort, while a sanction informs the player that his or her partner chose low effort. On the other hand, players that exerted low effort are sanctioned regardless of the behavior of their partner. Thus, sanctions are 'bad news' for someone who chose high effort in the first round and carry no news for those that chose low effort.

In summary, we can say that a sanction is never good news about the effort of the partner, and is bad news when own effort is high and the third player is believed to "sanction if at least one player exerted low effort". Careful inspection of the equilibrium behavior of agents leads to the following hypotheses.

Hypothesis 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. (c) On aggregate, the change in efforts and beliefs will be larger under exogenous sanctions than under endogenous sanctions.

Note that the model also predicts that effort among those not sanctioned will not change. However, we do not formally test this hypothesis as our primary interest is the effect of sanctions.

5. Results

The number of participants in the experiment was 243: 45 in the ExNS treatment, 51 in the ExS treatment, and 147 in the endogenous treatment where the principal decided to introduce a sanction in 29 out of 49 groups. Each experimental session lasted roughly 35 min and the subjects earned 7.5 euros on average.[3] As mentioned above, participants engaged in a 5-min tutorial before starting the experiment and being assigned to a role. As an indication of whether people understood the game, we 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. In fact, there is a high correlation between beliefs (as identified by the lower bound L of the elicited belief interval) and efforts in the first round of each treatment, as one would expect in a minimum-effort game. The correlation coefficient is 0.85, which is significant at the 1% level.[4]

Fig. 1 shows a histogram of first-round effort choices. Average effort in round one across all treatments was 145 with a large clustering of observations around 170 and a smaller cluster around 110. Further descriptive statistics are given in Appendix C.

Before we move to test the hypotheses formulated in Section 4, some comments on methodology are in order. First, we prefer not to add any unwarranted distributional or parametric assumptions, in particular as our sample sizes are small. Our main analysis is based on new nonparametric tests for 'stochastic inequality' that enable correct inference for the given sample sizes and that have been specifically designed for small samples (Schlag 2008).[5] The power of these tests stems from the fact that they are based on ordinal comparisons and hence are less sensitive to outliers, as is explained in more

[3] If this seems to be 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.

[4] The significance is based on an exact test of Schlag (2008) which has as null hypothesis that the covariance is less than 0. Note that this is not the null hypothesis underlying the Spearman rank correlation test (Spearman, 1904).

[5] Software for the implementation of these tests can be downloaded from http://homepage.univie.ac.at/karl.schlag/. Note that all our results are consistent with those that can be obtained using the Wilcoxon–Mann–Whitney (WMW) test. However, contrary to conventional wisdom, the WMW test is not an exact test for comparing means unless one is willing to assume that any two random variables that are not identically distributed have different means (for a counterexample, see Forsythe et al., 1994).

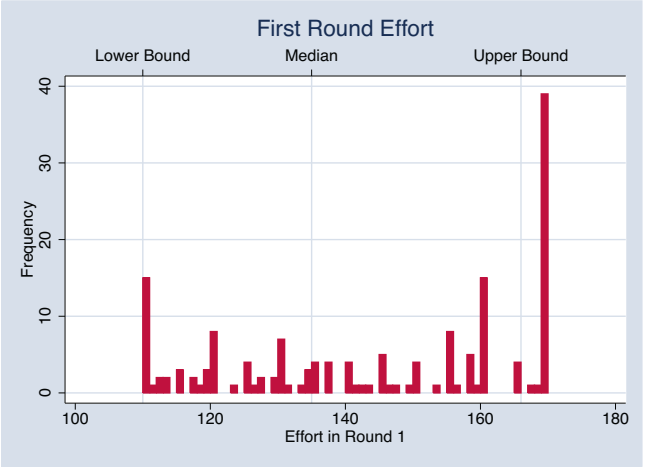

Fig. 1. Histogram of first-round effort choices.

detail in Appendix B.[6] Since these tests cannot account for potential endogeneity problems, we revisit our analysis within a parametric framework in Section 5.4. The results of the two different methodologies are consistent with each other.

Second, our hypotheses are framed in terms of changes in these variables between round one and two rather than absolute levels. Thus, our main (nonparametric) analysis is based on comparing differences. To indicate changes between the two rounds of a variable in treatment X we use the notation dX.

Third, interpreting changes in efforts and belief at the boundary of the interval is problematic. Our predictions are that subjects raise their effort in a reaction to the introduction of exogenous sanctions. However, people who chose effort very close to 170 in the first round will not be able to move up their effort any further. The behavior of these subjects is therefore not informative about the effect of sanctions. Furthermore, these individuals have a lot of room to move down, so that mean reversion may play an exaggerated role. Thus, we restrict attention to those subjects who can respond to incentives in either direction, and choose a first-round effort below an upper bound of 165 (indicated in Fig. 1) in our analysis of efforts and first round beliefs below 165 in our analysis of beliefs.[7] The results are robust to changes in this threshold.[8] Consequently, the number of observations of beliefs and effort may differ due to a different number of observations above the threshold.

Fourth, our hypotheses instruct us to condition on first-round effort, so we differentiate between subjects with high and low first-round effort. As cutoff between the two regions we consider the sample median effort of the remaining subjects in round one which is 135. Thus, in the remainder we define high effort players as those who chose effort in the first round in $\{135, \ldots, 165\}$ (i.e. above the median), and low effort players as those chose first-round effort in $\{110, \ldots, 134\}$ (i.e. below the median).

Finally, the observations for the group members in the endogenous treatment are not necessarily independent. The effort decision of a subject in the first round can influence the sanctioning decision of the third player. This in turn can influence the effort and beliefs of the other subject in the second round. In our statistical testing we correct for this dependence by considering the average effort/belief level within each group whenever the sample contains two players from the same group (which may not always be the case since we split the sample between high and low effort players).

5.1. The incentive effect of sanctions

In order to identify the incentive effect of sanctions (Question 1), we compare behavior in the exogenous no-sanction (ExNS) treatment to that in the exogenous sanction treatment (ExS). Fig. 2 shows the change in the mean belief and mean effort between round 1 and round 2 for both treatments. The number on top of the bar indicates the number of independent observations.

In the first column of Table 1 we report the estimated stochastic differences between the change in ExNS and the change in ExS with their respective significant levels. The estimated stochastic difference equals 0.64, and we are able to reject the

[6] Formally, given two random variables Y_1 and Y_2, the stochastic difference between Y_1 and Y_2 is given by $\delta(Y_1, Y_2) = \Pr(Y_2 > Y_1) - \Pr(Y_2 < Y_1)$. We thus test the null hypothesis that $\Pr(Y_2 > Y_1) \leq \Pr(Y_2 < Y_1)$. A rejection presents evidence that data drawn from Y_2 tends to be larger than data drawn from Y_1.

[7] Those with first-round effort near 110 also face a constraint, but this is less problematic since we hypothesize that people move up in reaction to incentives. In fact, subjects do not seem to be constrained. We find that no subjects with low effort (see below) in the first round decreased their effort and only 3 subjects who had low beliefs in the first round decreased their beliefs in round two.

[8] Our focus on subjects with first-round effort below 165 eliminates 39 effort observations and 11 belief observations, including two extreme outliers in the ExS treatment, with changes in effort equal to −60 and −51. The results of our non-parametric tests hold for any upper bound between 165 and 168, when we take the median of the associated sample as a threshold between high and low effort players. They do not hold if we include the spike of observations at 170, which results in a large number of no-change observations which swamp the statistical differences between treatments.

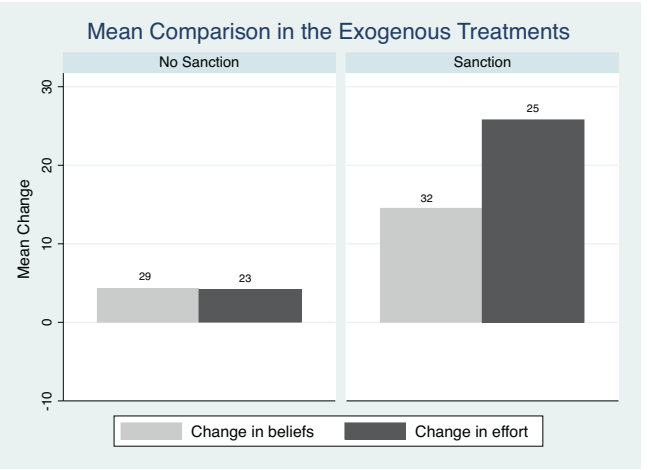

Fig. 2. The change in beliefs and sanctions for those who chose first-round efforts $\in\{110, 111, \ldots, 165\}$ or first round beliefs $\in\{110, 111, \ldots, 165\}$. The number of independent observations for each sample is at the top of the bar.

null hypothesis that the stochastic differences is nonpositive at a significance level of 1%. The impact of sanctions on the change in beliefs is only marginally significant (10%). Thus, it seems that sanctions have a stronger impact on efforts than they do on beliefs. Testing this formally, we find that the null hypothesis that the change is equal or higher for beliefs is rejected at the 10% level, with the stochastic difference being 0.35.

Summary 1. Regarding Hypothesis 1, we find a strongly significant incentive effect of sanctions on effort and a marginally significant belief effect. There is marginally significant evidence that the effect is stronger for effort than it is for beliefs.

Fig. 2 also suggests that average beliefs and efforts increase even in the absence of sanctions. The second column of Table 1 tests whether this effect is significant, where ExNS1 (ExNS2) denotes the first (second) round choices in the exogenous no-sanction treatment. Here, a test of the null hypothesis that stochastic difference is 0 can be performed with a sign test. We find no marginally significant difference for effort (the WMW test also does not detect any statistically significant difference at 10%). On the other hand we do find significant evidence that the beliefs tend to be higher in the second round. However, the changes in beliefs are sufficiently small that people do not change their effort levels by much.

5.2. The signaling effect of sanctions

The second main objective of this paper is to investigate the signaling effect of sanctions. To this end we study behavior in the endogenous treatment, using the exogenous treatments as controls. Signaling occurs when the principal conditions the choice of whether to sanction on the effort levels of the two players chosen in round one. Signaling has an effect when subjects make inferences about the effort of the other player when observing the choice of the principal whether or not to sanction. Note that the information contained in the choice of the principal need not be consistent with how subjects interpret why the principal chose to or not to sanction. Therefore, we separately analyze (i) the sanctioning choice of the principal and (ii) how subjects react to the choice of the principal.

5.2.1. The sanctioning decision

We wish to uncover regularities in the sanctioning choice of the principal. In particular, Hypothesis 2 predicts that sanctions are more likely when the minimal effort in round one is low. In order to test Hypothesis 2, we compare the minimum first-round effort in the sanctioned groups to the minimum first-round effort of non-sanctioned groups in the endogenous treatment.

Table 1
Column one shows values of stochastic difference and their significance level as defined in footnote 6 between *changes* in the ExNS treatment (dExNS) and *changes* in the ExS treatment (dExS). Similarly, Column 2 compares round 1 and round 2 in the ExNS treatment.

	Stochastic difference	
	dExNS vs. dExS	ExNS1 vs. ExNS2
Effort	0.64***	0.22
Belief	0.31*	0.34**

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

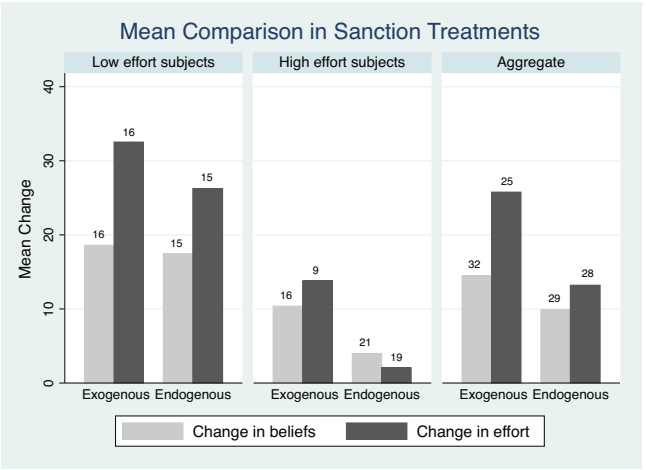

Fig. 3. Means of changes in beliefs and effort in the ExS and EnS treatment, for those who chose low effort ($\in\{110, 111, \ldots, 134\}$) in the first round (left panel), high effort/beliefs ($\in\{135, 136, \ldots, 165\}$) in the first round (middle panel) and the aggregation of those two samples (right panel). The number of independent observations for each sample is at the top of the bar.

When we compare the minimum group effort between the sanctioned and the non-sanctioned groups, we see that the group effort is slightly lower on average in sanctioned groups (135 vs. 138). However, using a Wilcoxon–Mann–Whitney (WMW) test, we do not find any significant evidence that the distributions of minimal effort are different between the two groups ($p = 0.63$). Since the samples are small, the test is not very powerful. A Probit regression of the probability to sanction yields a negative but insignificant coefficient for minimum effort ($p = 0.80$).

Summary 2. Although the minimum-effort group effort is lower on average in sanctioned groups, we find no statistically significant evidence in favor of Hypothesis 2. We cannot reject the hypothesis that sanctions have been imposed independently of the minimum-effort.

There may be several explanations why we cannot confirm Hypothesis 2. One explanation for the lack of pattern may be the established tendency of people to sanction "too often" in economic experiments (Fehr and Rockenbach, 2003). This would explain why we occasionally observe sanctions even if the minimum-effort was quite high. Note however that this result does not contradict our theoretical framework outlined in Appendix A, since this admits equilibria in which sanctions are chosen independent of effort levels.

5.2.2. The effect of endogenous sanctions

We now turn to investigate whether subjects perceive an informational content in the choice of the principal by comparing behavior in the ExS and EnS treatments. Note that there is a potential confound due to the endogenous nature of the EnS treatment. Even though we could not detect a significant relation between minimum effort in round one and the imposition of a sanction, we cannot rule out that there is such an effect. Moreover, the principals may have systematically used some other feature of first round effort in the sanctioning decision, a feature which may drive behavior in the second round. Here we ignore this endogeneity issue, since it is hard to solve with non-parametric tests, and come back to it in Section 5.4 where we address it using regression analysis.

Although our analysis above indicates that the actual informational content in the sanction is small, subjects may still believe that sanctions are imposed as a reaction to low minimum-effort levels. Specifically, subjects may follow the same reasoning that led us to formulate Hypothesis 2. If this is the case, sanctions will influence beliefs about the other group member, and the predicted effect depends on whether a subject chose low or high effort in round one. We now investigate separately the behavior of low and high effort players. Descriptive statistics for these samples can be found in Appendix C.

Low effort players. Consider the behavior of low effort players. The left panel of Fig. 3 presents the mean changes in beliefs and effort for people who chose low effort in the first round. It reveals no large differences between the exogenous and endogenous sanction treatments.

The first column in Table 2 shows that we cannot reject the null hypothesis of identical distributions in the exogenous and endogenous treatments, using the WMW test, both for effort and beliefs.

The fact that we are not able to reject the null hypothesis does not mean that there is no effect for low effort players. To obtain more evidence regarding Hypothesis 3a, we contrast the stochastic difference of the change in behavior between the two rounds of the two sanction treatments, reported in the second and third column of Table 2. Estimates and levels of significance are very similar.

Table 2
Comparison of exogenous and endogenous sanction treatment among those that chose low effort ($\in\{110, 111, \ldots, 134\}$) in the first round. First column: p-values of the two-sided WMW test for no-difference hypothesis. Second and third column: stochastic difference between the first and second round effort in the ExS and EnS treatment, respectively.

	WMW p-values	Stochastic difference	
	dEnS vs. dExS	ExS1 vs. ExS2	EnS1 vs. EnS2
Effort	0.29	1***	1***
Belief	0.97	0.83**	0.93***

*$p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 3
Comparison between the differences in the EnS and ExS treatments in the first round (first column), between the ExNS and EnS no sanction treatment (second column) and between the ExNS and ExS treatment (third column), for those who played high effort ($\in\{135, 136, \ldots, 165\}$).

	Stochastic difference dEnS vs. dExS	WMW p-value dExNS vs. dEnS	Stochastic difference dExNS vs. dExS
Effort	0.66**	0.35	0.78***
Beliefs	0.39*	0.49	0.48**

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 4
Estimates of stochastic difference between the EnS and ExS treatment and the EnS and ExNS treatment for those who chose first round effort $\in\{110, 111, \ldots, 165\}$.

	Stochastic difference	
	dEnS vs. dExS	dEnS vs. dExNS
Effort	0.4**	0.34*
Belief	0.13	0.21

* $p < 0.10$.
** $p < 0.05$.
***$p < 0.01$.

Summary 3. Regarding Hypothesis 3a, for subjects who chose low efforts in the first round we find no statistically significant evidence that endogenous and exogenous sanctions have different effects on either efforts or beliefs. There is some indication that lack of significant difference is not due to small sample sizes but that in fact behavior is the same under endogenous and exogenous sanctions.

High effort players. Consider now the behavior of high effort players. In the second panel of Fig. 3 we report average changes in efforts and beliefs across treatments for subjects who chose high effort and beliefs in the first round. In line with Hypothesis 3c, exogenous sanctions seem to raise effort and beliefs more than endogenous sanctions in this sample. Our statistical analysis based on stochastic differences, reported in the first column of Table 3, confirms this with respect to effort and (somewhat less significantly) beliefs.

One might wonder whether endogenous sanctions have any effect at all. To find out, we test if there is a difference between the EnS and the ExNS treatment. In the second column of Table 3 we report the p-values of the WMW test for this comparison. There is no statistically significant evidence that endogenous sanctions raise effort amongst high effort players (a test for stochastic difference is similarly insignificant). 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. The last column of Table 3 shows the stochastic difference when the sanction is imposed exogenously. Despite the small sample sizes, we find statistically significant evidence that exogenous sanctions are effective among the high effort players.

Summary 4. Regarding Hypothesis 3b, for subjects who chose high effort in the first round, endogenous sanctions are significantly 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.

Aggregate effect of endogenous sanctions. We now consider the effect of endogenous sanctions for both low and high effort players combined. In line with Hypothesis 3c, the right panel of Fig. 3 shows that the mean change of effort is almost twice as large under exogenous sanctions as it is under endogenous sanctions (25.8 vs. 13.6). The first column of Table 4 shows that exogenous sanctions tend to raise effort more than endogenous sanctions. The effect for beliefs goes in the same direction, but is not significant.

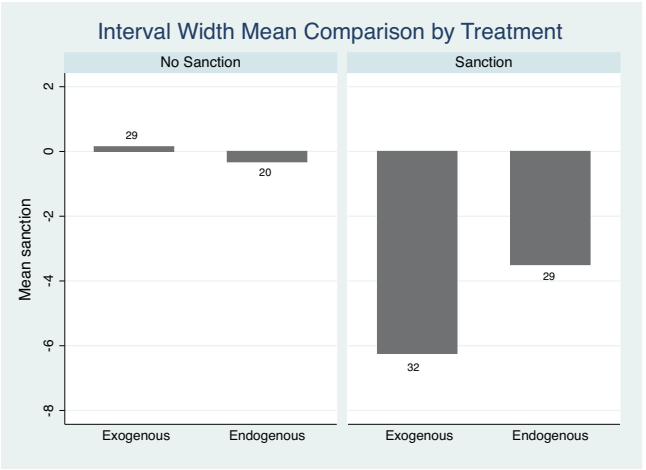

Fig. 4. Means of change in the width of the interval across treatments, for those who chose the lower belief interval in the first round in ($\in\{110, 111, \ldots, 165\}$) in the first round (number of independent observations for each sample at the top of the bar).

Relative to the no-sanction case, the second column of Table 4 shows marginally significant evidence that endogenous sanctions tend to raise effort. Changes in beliefs are not significantly different from the no-sanction case. Note that this evidence for the effectiveness of endogenous sanctions is much weaker than the corresponding evidence for exogenous sanctions reported in Table 1. Thus, as was to be expected, the aggregate results fall in between the results derived separately for the low and high effort players.

Summary 5. Regarding Hypothesis 3c, we find that endogenous sanctions are significantly less effective in raising efforts than exogenous sanctions, but do not find similar effects for beliefs.

5.3. Sanctions and uncertainty

One of the reasons we asked the participants to specify an interval rather than a point belief was that the elicited interval provides some indication of the uncertainty about the behavior of the other player. Schlag and van der Weele (2011) show formally that if subjects maximize expected utility, changes in the width of the interval are a proxy for changes in uncertainty. Fig. 4 shows the changes in the width of the belief interval for those who chose L in $\{110, 111, \ldots, 165\}$ in the first round. Uncertainty did not change between rounds in both no-sanction treatments, while uncertainty went down in both sanction treatments.

Statistical analysis confirms these results. In the no-sanction treatments a test of stochastic inequality cannot reject the null hypothesis that the distributions in the two rounds are equal at the 10% level. By contrast, we find that there is significant evidence (at the 5% level) that the interval decreases under exogenous sanctions and marginally significant evidence (at the 10% level) 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 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 a strongly significant decrease in uncertainty at 1% in the exogenous sanction treatment, while under endogenous sanctions it is no longer significant.

Summary 6. We find significant evidence that uncertainty about the choice of the other player is reduced when sanctions are imposed. This is also true within the subset of high effort players when the sanction is exogenous. No statistical evidence of a reduction in uncertainty is found in absence of sanctions or among high effort players when the sanction is introduced endogenously.

5.4. Discussion

In this section we present further evidence related to alternative interpretations of the results. We first ask whether the results could be driven by the principal's selection of whom to sanction, and then consider whether negative reciprocity towards the principal may play a role.

5.4.1. Endogeneity of sanctions

Inherent in our experimental design is that sanctions are implemented endogenously in some of the treatments and not in others. Although we could not reject the hypothesis that sanctions were implemented independently from the minimum

Table 5
Regression analysis of the effect of exogenous and endogenous sanctions on second round efforts and beliefs, assuming homoskedastic errors. All specifications include dummies for each level of first round effort. Sanction (or "S") is a dummy that is 1 in the EnS and ExS treatment and 0 otherwise, "Endo" indicates a dummy that is 1 for the EnS and EnNS treatment and 0 otherwise, "Effort1$_{[x,y]}$" is a dummy that takes the value of 1 if first round effort is in the interval $[x, y]$, and 0 otherwise. Standard errors (clustered by group in the endogenous treatment) in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	Effort R2	Beliefs R2	Effort R2	Beliefs R2	Effort R2	Beliefs R2
Belief Round 1	0.257	0.479**	0.0636	0.397*	0.0872	0.386*
	(0.227)	(0.199)	(0.211)	(0.202)	(0.224)	(0.213)
Interval width Round 1	0.149	−0.184	−0.0313	−0.286	−0.0275	−0.304
	(0.226)	(0.220)	(0.215)	(0.225)	(0.233)	(0.236)
Sanction	8.959*	8.712**	97.23***	46.34*		
	(4.742)	(3.509)	(27.02)	(24.76)		
S × Endo	−1.295	−1.267	−22.56	14.21		
	(4.964)	(4.014)	(36.62)	(31.14)		
S × Effort1			−0.605***	−0.254		
			(0.192)	(0.158)		
S × Endo × Effort1			0.157	−0.102		
			(0.243)	(0.195)		
S × Effort1$_{[110,134]}$					23.60***	16.94**
					(6.659)	(7.606)
S × Endo × Effort1$_{[110,134]}$					−0.407	2.340
					(9.032)	(8.925)
S × Effort1$_{[135,165]}$					13.69***	8.383**
					(5.208)	(3.619)
S × Endo × Effort1$_{[135,165]}$					−12.40***	−4.746
					(4.540)	(3.770)
S × Effort1$_{[166,170]}$					−6.020	3.622
					(9.109)	(4.350)
S × Endo × Effort1$_{[166,170]}$					9.771	−2.016
					(8.698)	(2.981)
Observations	162	162	162	162	162	162
R^2	0.482	0.595	0.554	0.604	0.556	0.607

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

effort in the first round, it may nevertheless be that the sample of subjects sanctioned by the principal is somehow different from the sample in the ExS treatment. A simple comparison finds that average first round effort levels are slightly higher in the EnS than in the ExS treatment (144.8 vs. 139.6). However, a WMW test does not reject the hypothesis that the two distributions are the same $p = 0.43$. Obviously, this does not prove that there is no endogeneity, but it does mean subjects in both treatments had similar first round effort levels.

Since nonparametric methods to address this endogeneity problem do not exist, we use parametric methods, acknowledging that these are based on assumptions that may not hold here.[9] We control for the possibility that differences in first-round choices may drive our treatment effects by running a series of OLS regressions, reported in Table 5.[10] The dependent variable is second round effort (odd columns) or beliefs (even columns). The dummy "Sanction" (or "S") is equal to 1 for the EnS and ExS treatment (and 0 otherwise), and "Endo" is 1 for the EnS and EnNS treatment (and 0 otherwise).[11]

To control for any effects that are due to first round choices, we include the interval width, first round beliefs as well as a series of 71 dummies for each level of first round effort (the coefficients are not reported for reasons of space). To account for the interdependence of observations, we cluster standard errors by group in the endogenous treatment.

The first two columns show that sanctions have a marginally significant effect on second round effort and a significant effect on second round beliefs, in line with Hypothesis 1. The additional effect of the endogeneity of sanctions is negative but not significant. When we control for the interaction of sanctions and first round effort in columns 3 and 4, we see a much stronger effect of sanctions on effort. The significant and negative coefficient on the interaction term indicates that higher efforts reduce second round effort in the sanction treatments, possibly because there is less space to move up. There is no significant difference in this effect between exogenous and endogenous sanctions.

In the last two columns we investigate the interaction of the sanction with a dummy if effort is either 'low' or 'high'. In line with Hypothesis 3, and consistent with the nonparametric analysis, for the low effort players we see a strong effect of sanctions and no significant negative effect of the endogeneity of sanctions. For the high effort players, we see a strong positive effect of sanctions which is almost entirely canceled out by a strong negative effect when sanctions are introduced

[9] In particular, the assumption of homoskedastic errors may fail, and it is not clear that that our sample size is large enough for asymptotic theory to justify the assumption of normally distributed errors.
[10] A series of Tobit regressions accounting for the fact that subjects were constrained to the interval [110, 170] yields very similar results.
[11] We thank an anonymous referee for helpful comments related to these models.

endogenously. Note that for effort in the interval [165, 170] there is no significant effect. The reversal of the signs for the endogenous and exogenous sanction is driven by two rather extreme outliers in the ExS treatment (see footnote 8), and does not occur for beliefs.

5.4.2. Can negative reciprocity explain the results?

There is an alternative potential explanation for the difference between the effect of endogenous and exogenous sanctions. Although the sanction does not operate retroactively, and the incentives of all players are aligned, subjects could nevertheless interpret the introduction of an endogenous sanction as an unkind act by the principal. If agents have reciprocal preferences, they may retaliate by reducing their effort in order to lower the payoffs of the principal.

If such retaliation drives the results, we should observe that subjects' effort choices deviate downward from their beliefs about the other player's effort in the EnS treatment.[12] One could test this in several ways. A non-parametric measure of punishment is the difference between effort and the lower bound of beliefs. If subjects wanted to punish the principal, we expect that more of them play efforts below the lower bound of the belief interval. In Fig. 5 in Appendix D, we show effort minus the lower bound of the belief interval in the second round for all treatments. The figure shows no clearly different pattern in the endogenous sanction treatment than in the other treatments.

A second test is to regress the change in effort on the change in beliefs, and dummies for treatment effects. If punishment plays a role, the endogenous sanction treatment should have an effect on effort changes that is not captured by the changes in beliefs. The results in Table 7 in Appendix D, show that the coefficient for the change in belief is significant at the 1% level, but the coefficients for the treatment effects are insignificant. This result indicates that all treatment effects go through the change in beliefs about the effort of the other player. Thus, we conclude that retaliation against the principal does not play an important role in our experiment.

6. Conclusion

The results of our experiment show that sanctions have a positive effect on effort levels and beliefs about others' effort level for those that chose low effort in the first round. For those that chose high effort their effect depends on whether sanctions were imposed exogenously or endogenously. Exogenous sanctions raise effort significantly, whereas endogenous sanctions do not have any effect.

These results can be explained by the signaling theory underlying our hypotheses. Subjects with high effort in round one interpret the sanction as a signal that their partner did not 'cooperate', i.e. she selected low effort. This explains why endogenous sanctions do not raise their beliefs about the effort of their partner in the next round. It also explains why for those with low effort in the first round, the effect of the sanction is independent of endogeneity. For them there is no signal, as the sanction would be imposed independently of their partner's behavior.

Comparing our experiment to the previous literature on endogenous sanctions, our experiment complements the results on democratically implemented sanctions mentioned in the introduction. This literature shows that introducing sanctions endogenously through voting raises expectations of cooperation and the effectiveness of such sanctions. Relative to this literature, our setup reflects more closely the arrangements of a society where laws are made by authorities, rather than by direct voting. In this setting, the superior information of the authority reverses the signaling effect, and sanctions are more effective when introduced *exogenously*. Note that the two effects are not mutually exclusive. In fact, in so far as the the decision to put a vote on the agenda in the first place is endogenous, our results suggest that even voting procedures may have adverse signaling effects.

Assuming some external validity, we believe our results have relevance for both public policies and manager-employee relationships in firms. For example, Brandts and Cooper (2006) stresses the relevance of coordination and the existence of inefficient equilibria for corporations and other organizations. The existence of signaling effects means authorities need to strike a balance between correcting the behavior of deviants or pessimists on the one hand, and maintaining the optimistic beliefs of cooperators on the other.

Finally, we hope to promote the use the elicitation of belief intervals, as well as statistical tests that are exact but do not, like the WMW test, impose additional distributional assumptions. We think these tests are an important addition to the toolbox of economists, in particular when working with small data sets.

Appendix A: A simple model and hypotheses

In this section we present a simple model of behavior for the game specified in Section 3.1. A summary is provided in Section 4.

We first predict behavior for the case where the third player does not have an active role. To generate predictions we make the following assumptions. Players believe that they are more sophisticated than the player they are matched with (whom we also call 'partner') and best respond to anticipated effort of their partner. Partners are believed to best respond to

[12] We thank Samuel Bowles for this suggestion.

a given belief distribution of effort levels where these beliefs do not change over time. Thus, the sophistication of players is as in the models of level k thinking or cognitive hierarchy (Nagel, 1995; Costa-Gomes and Crawford, 2006). In the terminology of these models, all players in our paper belong to level 2.[13]

Partners best respond to one of two different belief distributions G_h and G_l, where we assume that G_h that G_l is 'higher' than G_l (e.g. G_h stochastically dominates G_l). Accordingly some partners choose high while others choose low effort, and they are referred to as high types and low types, respectively. Each player assesses a probability or belief p that her partner is the high type. Players and partners are risk-neutral. Finally, players choose their effort levels in the first round as if there was no second round, so completely myopically. This is in accordance with the experimental setup, where people did not know in the first round that there would be a second round.

7.1. First-round effort

To determine her effort in the first round, a player will first calculate the optimal effort of the high type and of the low type partner and then choose a best response on the basis of the probability p of meeting the high type. Denote the optimal effort levels of the high and low type when there is no sanction by $e_h(0)$ and $e_l(0)$, respectively, where 0 indicates that there is no sanction (in later sections a 1 will indicate that a sanction has been imposed). So $e_h(0) \in \arg\max_e \left(\int \min\{e, e'\} dG_h(e') - ke \right)$ where the cost of effort k was equal to 0.85 in the experiment. We assume that G_h and G_l are such that $e_h(0) > e_l(0)$. According to our assumptions, each player believes with probability p that she faces a partner who chooses $e_h(0)$ and with probability $1-p$ a partner who chooses $e_l(0)$. Let e_p^r denote the optimal effort level of a player with belief p in round r, $r = 1, 2$. Note that $e_p^1 = e_l(0)$ if $p = 0$ and $e_p^1 = e_h(0)$ if $p = 1$. Taking into account that $e_p^1 \in [e_l(0), e_h(0)]$ holds for all $p \in [0, 1]$ we can write the expected utility of a player with belief p who exerts effort $e \in [e_l(0), e_h(0)]$ as $Eu = p(e - ke) + (1-p)(e_l(0) - ke)$ and obtain

$$\frac{d}{de} Eu = p - k.$$

So if $p > k$ then $e_p^1 = e_h(0)$, if $p < k$ then $e_p^1 = e_l(0)$.

7.2. The effect of an exogenous sanction

We now consider choice of effort in round 2 when an exogenous sanction has been imposed. Imposing a sanction means to subtract $k_1(170 - e)$ from the payoff for some given $k_1 > 0$. In the experiment we set $k_1 = 0.5$. This change in payoffs influences effort choices of the level 1 player. Let $e_v(1)$ be the optimal effort of type $v \in \{h, l\}$ when there is a sanction. So $e_h(1) \in \arg\max_e \left(\int \min\{e, e'\} dG_h(e') - ke - k_1(170 - e) \right)$. Note that $e_v(1) \geq e_v(0)$ for $v \in \{h, l\}$, i.e. partners (are believed to) exert more effort after a sanction has been imposed.

Expected utility of a player who exerts effort $e \geq e_l(1)$ is now

$$Eu = p(e - ke - k_1(170 - e)) + (1-p)(e_l(1) - ke - k_1(170 - e)).$$

Hence, $d/de Eu = p - (k - k_1)$. If $p > k - k_1$ then $e_p^2 = e_h(1)$, if $p < k - k_1$ then $e_p^2 = e_l(1)$. Thus, all players exert weakly more effort after an exogenous sanction has been introduced.

We can decompose this change in effort into two effects. First, there is an *incentive effect*, because a sanction effectively reduces the cost of effort k and thus gives incentives for higher effort. Specifically, any player with $p \in (k - k_1, k)$ chooses the effort of the high type in round 2 while they choose the effort of the low type in round 1. Players with $p < k - k_1$ and $p > k$ choose the same effort in round 2 as they do in round 1. Thus, taken over the whole sample, there will be a strict increase in average effort as long as $p \in (k - k_1, k)$ for at least some players. Second, there is a (forward looking) *belief effect* because introducing a sanction leads to a belief that partners will choose a higher effort as they too face lower effort costs. This belief effect additionally raises the effort levels of the players. This belief effect will be strictly positive as long as G_l and G_h place enough mass around the optimal choice.

We now compare the effect of a sanction on beliefs to their effect on effort levels. For this, we assume that players are drawn from some distribution such that G_p describes the distribution of p. To simplify exposition, assume that G_p has no point masses and full support on [0, 1]. Then the expected beliefs (in terms of the expected effort of a partner) in round one equals

$$\int (p e_h(0) + (1-p) e_l(0)) dG_p(p),$$

[13] At the cost of substantial additional complexity, one could assume more sophisticated distribution of rationality levels. Specifically, specifying higher levels of rationality would lead to more complex belief effects. We believe that the data do not justify the cost of such an analysis.

while the expected effort (of a player) in round one is equal to

$$\int e_p^1 dG_p(p) = G_p(k)e_l(0) + (1 - G_p(k))e_h(0).$$

In order to make efforts and beliefs comparable in round two we impose a mild consistency requirement, namely that expected beliefs equal expected effort in the first round. Following the above this means that $\int p dG_p(p) = 1 - G_p(k)$.

In round 2, invoking consistency, we find that expected beliefs equal

$$\int (pe_h(1) + (1 - p)e_l(1))dG_p(p) = G_p(k)e_l(1) + (1 - G_p(k))e_h(1),$$

and that expected effort equals

$$G_p(k - k_1)e_l(1) + (1 - G_p(k - k_1))e_h(1).$$

Comparing these two terms we conclude for round 2 that expected effort is higher than expected belief. Given that these two expressions are by assumption equal in round 1 we obtain the following result.[14]

Result 1. Exogenous sanctions increase both beliefs and effort where effort increases more than beliefs.

Note in the treatment where no sanction is introduced in round 2, payoffs and beliefs remain unchanged and hence $e_p^2 = e_p^1$, i.e. efforts remain unchanged as well.

7.3. The effect of an endogenous sanction

Next we investigate behavior when it is the third player, who we refer to as principal, who chooses whether or not to sanction. The principal's payoffs are given by $0.25 \min\{e_1, e_2\} - cs$ where in our experiment we set $c = 4$. Note that $c/0.25 = 4c$ is the cost of sanctioning in units of efforts. The principal is risk neutral and has a prior G_p over the possible values of belief p held by the players.

We develop some notation. Let $e_p(s)$ be the optimal effort given belief p where $s = 1$ ($s = 0$) indicates that a sanction has been imposed (has not been imposed). Let p_i be the belief of player i, $i = 1, 2$. Let $p_m = \min\{p_1, p_2\}$ and $p_x = \max\{p_1, p_2\}$. Let $s^* : [110, 170]^2 \to \{0, 1\}$ be such that $s^*(e_{p_1}^1, e_{p_2}^1)$ is the choice of the principal of whether or not to sanction conditional on observed effort level $e_{p_i}^1$ of player i in round 1, $i = 1, 2$.

Choices in the first round are assumed to be myopic as players do not anticipate that there will be a second round. We will not consider deviations from such play. Thus, the principal will observe only effort choices belonging to $\{e_l(0), e_h(0)\}^2$ and only needs to condition on these. We call $e_l(0)$ and $e_h(0)$ a low and a high first-round effort, respectively. We will consider only sanctioning strategies where sanctioning choices do not depend on player indices but only on effort levels. Thus we can identify $s^* : [110, 170]^2 \to \{0, 1\}$ with $s^* \in \{0, 1\}^3$ where s_1^*, s_2^* and s_3^* are the sanctioning choices conditional on the first round events $\{(e_l(0), e_l(0))\}$, $\{(e_l(0), e_h(0)), (e_h(0), e_l(0))\}$ and $\{(e_h(0), e_h(0))\}$, respectively.

We will make predictions that satisfy the following requirements.

1 The strategies of the principal and the two players can be supported as a Perfect Bayesian Equilibrium (PBE). Out of equilibrium actions of the principal do not change the belief of a player about her partner's effort.
2 The PBE does not depend on the specific form of the prior of the principal.
3 The PBE can be sustained for a non-degenerate interval of values of c.

We make some comments before we turn to the analysis. Given the assumptions above, a PBE is uniquely characterized by the sanctioning function $s^*(,)$ of the principal. Following requirement 2 the equilibrium candidate must be optimal, regardless of the beliefs over p_1 and p_2. This implies that it will be sufficient to evaluate deviations from an equilibrium candidate using a degenerate prior of the principal, i.e. when the principal is (almost) sure about p_1 and p_2. If the principal does not want to deviate under any degenerate prior, she will also not want to do so under more general priors. To see this, it suffices to note that expected payoffs of a deviation under a general prior are just a convex combination of payoffs under some degenerate priors, and therefore cannot be strictly higher.

There are $2^3 = 8$ candidates for a PBE. In two of these the principal's choices are unconditional: $s^* = (1, 1, 1)$ and $s^* = (0, 0, 0)$. To "always sanction", i.e. $s^* = (1, 1, 1)$, can be supported if and only if $e_{p_m}(1) - 4c \geq e_{p_m}(0)$ holds for all p_m. Here we use our requirement that beliefs p_i do not change when the principal chooses the out of equilibrium action to not sanction. Necessary

[14] Without the consistency requirement, this is not necessarily true. As a counter-example, consider the case where high type partners have point beliefs and hence do not respond to lower effort costs. Assume furthermore that beliefs are such that both players choose high effort in the first round. As the effort of the high type partner remains unchanged in round two, players' effort remains unchanged too. Yet if some probability is put on the low type partners and if these respond to changes in effort cost, we find that beliefs move more than effort. However, this scenario occurs only if beliefs are inconsistent in the sense that first-round efforts are higher than first-round beliefs.

and sufficient conditions are given by $e_h(1) - 4c \geq e_h(0)$ and $e_l(1) - 4c \geq e_l(0)$. When investigating "never sanction", i.e. $s^* = (0, 0, 0)$, special attention must be given to a player with $p \in (k - k_1, k)$. A sanction would induce this player to switch from low to high effort, this is not in the interest of the principal if $e_h(1) - 4c \leq e_l(0)$. In fact, this is a necessary and sufficient condition for supporting "never sanction".

An intuitive conditional strategy is given by $s^* = (1, 1, 0)$ where the principal sanctions if and only if at least one of the two players chose a low effort in the first round. The conditions supporting this as a PBE emerge when considering three subcases. When both players chose low first-round effort then s^* prescribes to sanction is best when $e_{p_m}(1) - 4c \geq e_l(0)$ holds for all p_m, hence when $e_l(1) - 4c \geq e_l(0)$. When both exerted high effort in the first round, then $s^* = 0$ which is best if $e_h(0) \geq e_l(1) - 4c$. Finally, consider the case where one player had a low and the other a high first round effort. Then $s^* = 1$ which yields outcome $e_l(1) - 4c$ as the player with high first-round effort now believes that her partner is of low type. Not sanctioning causes the player with low first round effort to choose $e_{p_i}(0) = e_l(0)$ and the one with high first-round effort to choose $e_h(0)$, which is worse if $e_l(1) - 4c \geq e_l(0)$. Together this means that $s^* = (1, 1, 0)$ can be supported if and only if $e_h(0) \geq e_l(1) - 4c \geq e_l(0)$. Note that in this equilibrium, sanctions are "bad news" in the sense that playing $s^* = 1$ will alert a high effort player to the fact that her partner chose low effort.

The five remaining strategies can all be ruled out by our requirements 1–3. It is easy to show that $s^* = (0, 1, 0)$, $s^* = (0, 1, 1)$, $s^* = (0, 0, 1)$ cannot be supported at PBE. Moreover, one can rule out $s^* = (1, 0, 0)$ and $s^* = (1, 0, 1)$ using requirement 3. We summarize as follows.

Proposition 1. *The following values of s^* are the only ones that can be supported as a PBE for all G_p for a nondegenerate set of c: (i) $(1, 1, 1)$ if $e_v(1) - 4c > e_v(0)$ for $v \in \{l, h\}$, (ii) $(1, 1, 0)$ if $e_h(0) > e_l(1) - 4c > e_l(0)$, (iii) $(0, 0, 0)$ if $e_h(1) - 4c < e_l(0)$.*

Proposition 1 implies that there is no unique prediction for whether or not the principal will sanction low types or whether or not she will sanction high types. With respect to players 1 and 2, their efforts remain unchanged if there is no sanction. Players with low first round effort who are sanctioned increase effort in the same way as under an exogenous sanction, because the sanctions do not change the belief about the type of player she is facing. However, the predicted change in effort of a player with high first-round effort is ambiguous. She will increase effort in case (i), but when sanctions are "bad news" as in case (ii), she may reduce effort.

Appendix B: Stochastic difference and inequality

In the following we present two tests for making ordinal comparisons between two random variables, one for matched pairs and one for independent samples. Both tests are invariant to monotonic transformations. For matched pairs we show how one can transform the data, to then perform a test for comparing two Bernoulli random variables. For comparisons based on two independent samples we review a new test of Schlag (2008). Both tests are designed to uncover how two distributions in small samples differ without adding distributional assumptions. Permutation tests such as the Wilcoxon and the Wilcoxon–Mann–Whitney test can only establish significant evidence that two distribtutions differ (for counter examples in independent samples see Forsythe et al., 1994).

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* between Y_1 verses Y_2 (Cliff, 1993). 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 in which $Y_1 \neq Y_2$ 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. Following Vargha and Delaney (1998) we call this a test of *stochastic inequality* (see also Brunner and Munzel, 2000).

For *matched pairs* as given by n independent observations of (Y_1, Y_2) one can proceed as follows. Replace events $Y_1 > Y_2$ by $(1, 0)$, $Y_1 = Y_2$ by $(0, 0)$ and $Y_1 < Y_2$ by $(0, 1)$ and apply the z-test for matched pairs (Suissa and Shuster, 1991).

For the case of *independent pairs*, the test for stochastic inequality by Schlag (2008) proceeds as follows. In a first step observations from two independents are randomly matched. One treats these as matched pairs and then proceeds to test as if these matched pairs are exogenously given. In a second step one controls for the randomness that is implicit in the matching. For more details see Schlag (2008).

Appendix C: Full sample descriptive statistics

See Table 6.

Appendix D: Testing reciprocity

See Fig. 5 and Table 7.

Table 6
Mean effort and belief levels (lower bound (L) and upper bound (U)) for the entire sample, as well as for those who played low effort ($\in\{110, 111, \ldots, 134\}$) and high effort ($\in\{135, 136, \ldots, 165\}$) players. For the high effort players the number of observations for beliefs may exert that of effort, since some players chose effort > 165≥ beliefs (L). Note that for the endogenous treatment the number of observations may not correspond to those in the figures and the statistical tests, since we use group means to correct for the potential interdependence of observations.

		# Obs	First round			Second round		
		Effort (Belief)	Effort	Belief (L)	Belief (U)	Effort	Belief (L)	Belief (U)
All	ExNS	30	141.8	134.9	150.9	144.0	139.0	155.2
	ExS	34	139.6	133.9	156.7	155.3	147.5	164.5
	EnNS	40	150.5	145.0	160.0	150.2	144.6	160.5
	EnS	58	144.8	141.3	158.1	155.9	150	164.2
Low effort	ExNS	10	119.2	115.2	130.2	132.8	129.5	147.6
	ExS	16	118.6	122.8	147.6	151.1	141.1	162.4
	EnNS	12	122.4	124.3	143.9	128.6	124.8	147.5
	EnS	21	119.2	122.4	144.4	147.3	140.8	159.9
High effort	ExNS	13 (19)	143.9	143.2	160.8	140.8	142.4	158.4
	ExS	9 (16)	146.7	140.6	164.2	160.4	150.9	165.8
	EnNS	14 (22)	155.4	149.8	166.1	154.8	150.1	164.9
	EnS	26 (32)	155.1	149.3	165.3	157.0	153.2	166.0

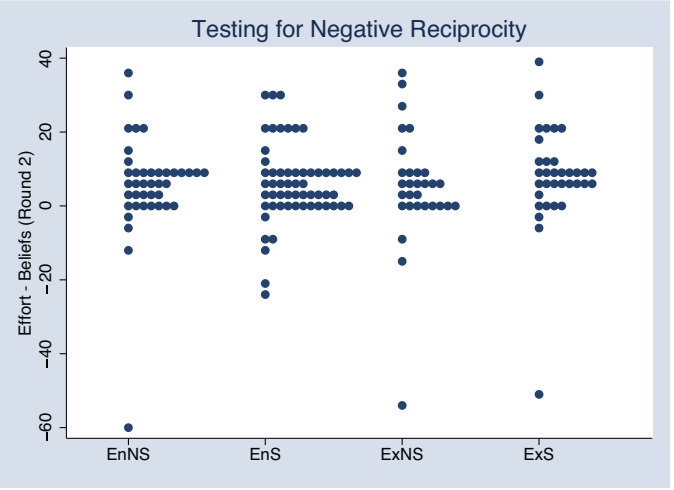

Fig. 5. Second round effort minus the belief lower bound for the full sample in all treatments.

Table 7
Regression of changes in effort on changes in beliefs, a dummy for the introduction of a sanction (takes a value of 1 for the ExS and EnS treatments) and for the EnS treatment, using the full sample. We assume homoskedastic errors. *t* statistics in parentheses.

	Effort change
Belief change	0.839***
	(10.53)
Sanction	4.759
	(1.51)
Endogenous sanction	−0.439
	(−0.14)
Constant	−0.514
	(−0.30)
Observations	162
R^2	0.472

*$p < 0.10$
**$p < 0.05$.
*** $p < 0.01$.

Appendix E. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2013.08.002.

References

Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Rev. Econ. Stud. 70, 489–520.
Bénabou, R., Tirole, J., 2011. Laws and norms. In: NBER Working Paper 17579.
Bowles, S., 2008. Policies designed for self-interested citizens may undermine the moral sentiments: evidence from economic experiments. Science 320.
Brandts, J., Cooper, D., 2006. A change would do you good: an experimental study on how to overcome coordination failure in organizations. Am. Econ. Rev. 96 (3), 669–693.
Bremzen, A., Khoklova, E., Suvorov, A., van de Ven, J., 2011. Bad News: An Experimental Study on the Informational Effects of Rewards. Amsterdam University, mimeo.
Brunner, E., Munzel, U., 2000. The nonparametric Behrens–Fisher problem: asymptotic theory and a small-sample approximation. Biometr. J. 42, 17–25.
Cliff, N., 1993. Dominance statistics: ordinal analyses to answer ordinal questions. Psychol. Bull. 114, 494–509.
Costa-Gomes, M., Crawford, V., 2006. Cognition and behavior in two-person guessing games: an experimental study. Am. Econ. Rev. 96, 1737–1768.
Devetag, G., Ortmann, A., 2007. When and why? A critical survey on coordination failure in the laboratory. Exp. Econ. 10, 331–344.
Fehr, E., Rockenbach, B., 2003. Detrimental effects of sanctions on human altruism. Nature 422, 137–140.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games Econ. Behav. 6, 347–369.
Frey, B.S., Jegen, R., 2001. Motivation crowding theory. J. Econ. Surv. 15 (5), 589–611.
Frey, B.S., Oberholzer-Gee, F., 1997. The cost of price incentives: an empirical analysis of motivation crowding-out. Am. Econ. Rev. 87 (3), 746–755.
Friebel, G., Schnedler, W., 2011. Team governance: empowerment or hierarchical control. J. Econ. Behav. Org. 78, 1–13.
Gneezy, U., Rustichini, A., 2000. A fine is a price. J. Legal Stud. 29, 1–17.
Goeree, J., Holt, C.A., 2005. An experimental study of costly coordination. Games Econ. Behav. 51, 349–364.
Goeree, J., Holt, C.A., 2001. Then little treasures of game theory and ten intuitive contradictions. Am. Econ. Rev. 91 (5), 1402–1422.
Kamei, K., 2010. Democracy and Resilient Pro-Social Behavioral Change: An Experimental Study. Brown University Department of Economics (available at: www.econ.brown.edu/students/kenju_kamei/JMP.pdf).
Markussen, T., Putterman, L., Tyran, J.-R., 2011. Self-organization for collective action: an experimental study of voting on formal, informal, and no sanction regimes. In: Working Paper 2011-4. Department of Economics, Brown University.
Nagel, R., 1995. Unraveling in guessing games: an experimental study. Am. Econ. Rev. 85 (5), 1313–1326.
Schlag, K.H., 2008. A new method for constructing exact tests without making any assumptions. In: Working Paper 1109. Department of Economics and Business, Universitat Pompeu Fabra.
Schlag, K.H., van der Weele, J.J., 2011. Incentives for Interval Elicitation. Manuscript, University of Vienna.
Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. Am. Econ. Rev. 97 (3), 999–1012.
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15, 72–101.
Suissa, S., Shuster, J.J., 1991. The 2×2 matched-pairs trial: exact unconditional design and analysis. Biometrics 47, 361–372.
Tyran, J.-R., Feld, L.P., 2006. Achieving compliance when legal sanctions are non-deterrent. Scand. J. Econ. 108 (1), 135–156.
Van der Weele, J.J., 2012. The signaling power of sanctions in social dilemmas. J. Law Econ. Org. 28 (1), 103–125.
Vargha, A., Delaney, H.D., 1998. The Kruskal–Wallis test and stochastic homogeneity. J. Educ. Behav. Stat. 23 (2), 170–192.
Xiao, E., 2013. Profit seeking punishment corrupts norm obedience. Games Econ. Behav. 77, 321–344.