Threat and Punishment in Public Good Experiments

David Masclet, Charles N. Noussair, and Marie Claire Villeval

October 12, 2009

Abstract: Experimental studies on social dilemmas have shown that while the existence of a sanctioning institution improves cooperation within groups, it has also a detrimental impact on efficiency. Could pre-play threats of punishment have the same beneficial impact on cooperation than sanctions without reducing efficiency? In our Threat treatment players can assign non-binding threat points of sanctions for each possible level of contribution before deciding on their contribution level. After learning the others' contributions, they choose how many points of sanction they actually assign to each of the other group members. We find that threats increase significantly the level of contributions but do not improve efficiency. In our Second Order treatment we introduce the possibility to sanction also deviations between threats and actual sanctions. This leads to lower threats and therefore to less cooperation.

Keywords: Threats, non binding announcements, sanctions, public good experiment
JEL Classifications: C92, H41, D63

Contact: David Masclet, CNRS, CREM, 7 Place Hoche, 35065 Rennes, France. Email: david.masclet@univ-rennes1.fr. Charles N. Noussair, Department of Economics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: C.N.Noussair@uvt.nl. Marie Claire Villeval, University of Lyon, CNRS, GATE, 93, Chemin des Mouilles, 69130, Ecully, France, IZA, Bonn, and CCP, Aarhus. E-mail: villeval@gate.cnrs.fr.

Acknowledgements
We thank participants at the 2009 International Meetings of the ESA in Washington, USA, and Innsbrück, the 2009 APET workshop on Behavioral Public Economics in Lyon, France, for constructive and helpful comments. We thank Elven Priour for programming and research assistance. Financial support from the *Agence Nationale de la* Recherche (ANR-08-JCJC-0105-01, "CONFLICT" project) is gratefully acknowledged.

1. INTRODUCTION

A large number of experimental studies have explored the conflict between individual behavior and collective interest in social dilemmas. In a Voluntary Contribution Mechanism game, each group member receives an initial endowment that she may allocate between her private account and a group account that returns a payoff to each individual. Each individual has a dominant strategy to invest all her endowment into the private account and none to the group account, whereas the highest total group payoff would be reached if all members would invest their entire endowment in the group account. Laboratory experiments have shown substantial contribution levels in the initial periods and a decay of contribution as the game is repeated (Marwell and Ames, 1979; Isaac *et al*., 1985; Andreoni, 1988; Isaac and Walker, 1988a; Ledyard, 1995). Among factors that influence contributions in the VCM context, the positive effect of communication –especially face-to-face (see Sally, 1995)- and pre-play announcement of contribution on outcomes has been first emphasized (Dawes et al, 1977; Isaac *et al*., 1985; Isaac and Walker, 1988b, 1991; Ostrom *et al*., 1992; Kerr and Kaufman-Gilliland, 1994; Krishnamurthy, 2001; Brosig *et al*., 2003). More recently, the attention has shifted to the study of the impact of punishment institutions on cooperation (Fehr and Gächter, 2000, 2002; Masclet *et al*., 2003; Noussair and Tucker, 2005; Bochet *et al*., 2006; Sefton *et al*., 2007; Carpenter, 2007a,b; Egas and Riedl, 2008). These studies have revealed the strength of altruistic punishment even among unrelated individuals. While the availability of a sanctioning opportunity improves cooperation, this institution exerts, however, a detrimental effect on efficiency. Indeed, costly sanctions destroy resources and may even backfire by generating counter-punishment (Fehr and Rockenbach, 2003; Houser *et al*., 2008; Nikiforakis, 2008; Denant-Boemont *et al*., 2007).

In these experiments, the threat of sanctions is likely to discipline group members (see for example Cinyabuguma *et al*., 2005) but these sanctions are not made explicit. Could non-binding but explicit threats of sanctions be sufficient to improve cooperation within groups, therefore allowing a reduction of the use of sanctions and their associated cost? Indeed, if the announcement of the intention to punish free riders increases the perceived punishment risk, the players might increase their contribution such that less sanctions are actually enforced. Both cooperation and efficiency would therefore be improved at the same time. On the contrary, does the introduction of explicit threats crowd-out the intrinsic motivation to cooperate by creating a hostile environment? The potential role of announcing intentions to sanction has been, however, left almost unaddressed in the experimental literature, although threats usually preexist to sanctions in real life.[1] Many examples could illustrate this statement, from the education of children to the management of international conflicts. We are only aware of three studies analyzing the behavioral impact of pre-announced endogenous threats of sanctions but in a different context (Dickinson and Villeval, 2008; Bochet and Puterman, 2009; Li *et al*., 2009).[2]

[1] This remark concerns explicit endogenous non-binding threats. The situation is somewhat different if one considers exogenous legal threats. There are a few papers studying the impact of legal threat campaigns on tax compliance behavior (for a recent example, see Fellner *et al.,* 2009).

[2] In a principal-agent experiment, Dickinson and Villeval (2008) allow the principal to announce threats of monitoring and sanctions and observe both a dominant disciplining effect of threats on effort and some crowding-out effect of threats. Li *et al*. (2009) introduce in a trust game threats of sanctions by the trustor before the trustee makes his return decision. They find that trustees reciprocate less when they face sanction threats. In public good games with sanctions, Bochet and Puterman (2009) allow people to make non-binding announcements about their possible contribution; after viewing the others' announcements, they could announce threats of punishing false announcements and promises. They find that in response to threats people who announced initially low contributions increased their announced contributions. In our

In this study we investigate experimentally the influence of non-binding pre-play announcements of sanctions on the level of contributions, on the actual implementation of sanctions, and on efficiency. In the first stage of a three-stage VCM game with groups interacting repeatedly, players choose how many non-binding points of sanction they threat to assign for each possible contribution level of each group member in the last stage of the game. In the second stage, after being informed on the total number of threat points announced by the other group members for each contribution level, each player decides how much to contribute to the group account. Then, after being informed on each of his group members' contribution level, he decides how many points of sanction to assign eventually. In contrast with threat points that are free, sanction points are costly to both the punisher and his target. Since threats are non-binding and non-credible, any profile of threats should be compatible with a subgame perfect equilibrium where all players always contribute zero and never punish. One might conjecture, however, that introducing such announcements may influence cooperation positively and limit the actual use of punishment. Their effectiveness may nevertheless decrease if threats are never followed by sanctions.

We also investigate in a second experimental treatment whether the possibility to observe individual differences between pre-play announcements and actual punishment affects the assignment of threats and their efficiency. The individuals who are willing to establish some credibility to threats in their group (although from a game-theoretic point of view, they should never be considered as credible) may be willing to sanction those group

game, people do not announce their contributions, they only announce threats of sanctions for each possible contribution of others.

members who do not enforce their threats. In this treatment, at the end of the first three stages each individual can observe both the threats and actual sanctions assigned by each of their group members, except those directed to him, and he can assign a second round of sanctions.[3] Do people anticipate that any deviation between pre-play announcements and actual punishment will be sanctioned in stage four and adjust consequently their punishment upward in stage three to make their punishment decision fit with their announcement? Or do they adjust their threat decisions downward? While the first effect should favor cooperation, this is obviously not the case with the second effect.

The experimental design consists of three different treatments. The Baseline treatment is almost identical to the two-stage game in Fehr and Gächter (2000). In the first stage, individuals have to decide on how many ECU (*Experimental Currency Units*) to contribute to the group account. In a second stage, players observe the individual contribution of each group member and can assign punishment points to any of their group members. The Threat treatment is similar to the Baseline except that a preliminary stage is included before contribution decisions. In this additional stage, players have to pre-announce a number of non-binding threat points for each possible contribution level indicating their willingness to sanction. These threat points are costless. At the end of this stage each player is informed on the total number of threat points assigned by the rest of the group for each possible contribution level. Our Second Order treatment replicates the Threat treatment except that in a fourth stage players are informed on each of their group members' threats and actual sanctions directed toward each other group member.

[3] This design differs from previous VCM studies with second order punishment (Cyniabuguma *et al*., 2006; Nikiforakis, 2008; Denant-Boemont *et al*., 2007) as we display information not only on individual punishment behavior but also on initial threatening behavior.

They are therefore able to measure the potential individual differences between threats and sanctions. In this final stage, players can assign additional punishment points. This treatment allows us to measure whether second-order punishment is directed towards those who do not implement their threats. It also enables to evaluate the consequences of a better correspondence between threats and sanctions, if any, on the level of punishment and on cooperation. Finally, we test the robustness of our results by varying the relative cost of sanctions in each of the three treatments. While in the main treatments one punishment point costs twice as much to the punishee than to the punisher, in the robustness tests, the cost of sanctions is symmetric.

We first find that contribution levels are significantly higher when threats are allowed as pre-announcements increase the perceived risk of punishment without inducing a crowding-out of motivation to contribute. Indeed, threat decisions are strong predictors of subsequent sanction decisions. Second, while threats succeed in improving cooperation within groups they fail improving efficiency. Indeed, after setting their threatening schedule players tend to punish more a same contribution than in the Baseline treatment. Therefore, the total amount of sanctions is not decreased by the introduction of pre-announcements. Third, allowing observability and punishment of individual differences between threats and actual sanctions induces less cooperation. Indeed, people reduce the difference between threats and sanctions by assigning significantly less threat points under the second order treatment than under the threat treatment to avoid second-order sanctions. The main results are robust to a change in the cost ratio of sanctions. The severity of threats is not affected by such a change but the effects of threats are, however, less persistent over time in the low-cost than in the high-cost condition. In addition, consistent with previous studies, our data indicate that cooperation is lower when the monetary consequences of punishment are lower.

The remainder of the paper is organized as follows. In section 2, we describe the experimental design and the protocol. Section 3 presents the results of the experiment. Section 4 discusses these results and concludes.

2. THE EXPERIMENT

2.1. Overview

Our experiment consists of three treatments. The Baseline treatment is close to Fehr and Gächter (2000). The game is two-staged. At the beginning of each period, each member of a group of four players receives an endowment of 20 ECU to allocate between a private account and a public account that yields 0.4 ECU to each member of the group for each ECU allocated to the group account by any group member. The more ECU are allocated to the group account, the lower her own but the greater the group's total earnings. At the end of the first stage, each participant is informed of her first-stage payoff, π_i^1, which writes:

$$\pi_i^1 = (20 - c_i) + 0.4\sum_{j=1}^{4} c_j \qquad (1)$$

where c_i is player i' s contribution to the group account. At the beginning of stage two, each player is informed on the total contribution of the group as well as on the individual contribution of each of the three other group members. Then, she has an opportunity to assign costly punishment points to each of the other members of their group. To avoid reputation effects across periods, participants were associated with a letter of the alphabet, A,..,D that was randomly changed after each period, which makes it impossible to establish a link between individual contributions or punishing decisions across periods. Each player can assign a certain number of punishment points to each other group member in the range from 0 to 10. Each point assigned costs one ECU to the punisher and two ECU to his target. Therefore player *i*'s payoff after the second stage is given by:

$$\pi_i^2 = \pi_i^1 - 2\sum_{j\neq i} p_j^{i2} - \sum_{j\neq i} p_i^{j2} \qquad (2)$$

where p_i^{j2} is the number of points assigned by *i* to *j* in the second stage, and p_j^{i2} the cost of receiving punishment from player *j*.

The Threat treatment is identical to the Baseline except that a preliminary stage was included at the beginning of the game. In this additional stage, the players were required to announce a hypothetical punishment level in the range of 0 to 10 for each possible contribution level of any group member (i.e., from 0 to 20). All group members' announcements were made simultaneously. Participants were also informed that this announcement was non-binding. Let us call these points 'threat points' to avoid any confusion with the actual punishment points distributed in the last stage of the game. In the second stage of the Threat treatment, each participant decides how much to allocate to the group account. This stage is identical to the first stage of the Baseline treatment with the notable exception that before contributing, the players are informed on the cumulated hypothetical punishment level announced by the three other group members. Precisely each participant is informed on the total number of threat points assigned by the three other members of his group for each possible contribution level. The third stage of the

Threat treatment is identical to the second stage of the Baseline. Each participant observes the individual contribution of others and can assign punishment points. Note that it is common information that the number of punishment points assigned by a player is not required to match the number of threat points he announced in stage one. The payoff function in this treatment is therefore the same as in the Baseline treatment.

The Second Order treatment replicates the Threat treatment except that another sanctioning stage is added after stage three. In stage four, each player is informed on the number of threat points and punishment points directed by each other player toward each player other than herself, so that she can observe any difference between the threats announced in stage one and the actual punishment assigned in stage three. Then, each player can assign additional punishment points, namely $p_i^{j,4}$. Note that to prevent direct revenge effects, individuals were never informed about who sanctioned them personally and by how much.[4] That is, player i observes p_j^{j4}, for all $j \neq i$, but not for $j=i$. The cost of these points is the same as for punishment points assigned in stage three. Therefore, the final payoff for individual i in this treatment writes:

$$\pi_i^2 = \pi_i^1 - 2\left(\sum_{j \neq i} p_j^{i3} + \sum_{j \neq i} p_j^{i4}\right) - \left(\sum_{j \neq i} p_i^{j3} + \sum_{j \neq i} p_i^{j4}\right) \qquad (3)$$

The three treatments have been run under two different conditions: a low and a high cost condition. Precisely, in the high cost condition, as explained above, each punishment point assigned costs one ECU to the punisher and reduces the target's payoff by two

[4] Nikiforakis (2008) reports an experiment where players can observe individual punishment behavior, which makes reprisals possible. He finds that the existence of a reprisal opportunity tends to offset the positive effect of punishment. Other studies have investigated the effect of allowing subjects to punish second order free riding (i.e. punishing those who failed to punish low contributors to the group account) (Cinyabuguma *et al.*, 2006, Denant-Boemont *et al.*, 2007). These experiments suggest that allowing sanction enforcement increases contributions.

ECU. In the low cost condition, each punishment point assigned has the same monetary cost of one ECU for the punisher and the target. Therefore, the payoff functions in the low-cost condition are the same as in the high-cost condition except that the multiplier of the second term of equations 2 and 3 is dropped. This additional condition constitutes a robustness test to changes in the parameters of the game.

2.2. Procedures

16 sessions have been conducted at the LABEX of the Center for Research in Economics and Management (CREM), University of Rennes I, France. Between 8 and 20 subjects participated in each session. Overall 200 participants were recruited from undergraduate courses and no subject participated in more than one session. The experiment was computerized using the Ztree program developed at the University of Zurich (Fischbacher, 2007). Participants interacted during 20 periods under a partner matching protocol. Table 1 summarizes information about the sessions.

[Table 1 about here]

2.3. Theoretical predictions and behavioral conjectures

In each treatment, the only subgame perfect equilibrium of the game, whether it is played once or finitely repeated, is for all players to always contribute zero to the public good and to never punish in any sanctioning stage. Since announcement is non-binding and therefore non-credible, any profile of threat level in the Threat and Second Order treatments is compatible with this subgame perfect equilibrium.

One might however conjecture that introducing threat opportunities may favor cooperation by inciting both the senders and the receivers of threat points to contribute more. Indeed, although non credible from a game-theoretical point of view, these threats could be taken seriously by the players as unstructured communication on contributions has been shown to increase contributions (Dawes *et al*., 1977; Marwell and Ames, 1979; Isaac and Walker, 1988a; Ostrom et al, 1992; Duffy and Feltovich, 2006). The reason behind this is that the announcements may be perceived as a signal of a future decision to punish and receivers may therefore condition their actions on this signal. As a consequence, our conjecture is that threats lead to more cooperation, less punishment and thus higher earnings. If this conjecture was verified, the society would benefit from higher contribution levels without the detrimental cost of punishment.

Turning next to the Second Order treatment, the expected behavioral effects of allowing the participants both to observe the difference between the threat points distributed and the punishment points actually assigned and to punish again in stage four are not clear cut. People may sanction the differences between the announced and the actual punishment. If so, we expect that this type of behavior should either reduce the intensity of threats or increase the severity of punishment to adjust threats and sanctions. The first option is more likely since it is less costly than the second option. But people may have other motives to punish in the last stage of this treatment, which could entail different consequences. For example, people may be willing to strengthen the sanctions of low contributors if they consider that the sanctions assigned in stage three are not high enough. Some may punish those who failed to sanction low contributors in stage three (indeed, sanction enforcement can raise second order free-riding, as studied by Yamagishi, 1986). This could reinforce cooperation in the next periods. Last, stage four-punishment may aim at counter-punishing blindly for having been punished in stage three, which should impact cooperation negatively in further periods. The final impact of these various possible motives is left to empirical evidence.

3. RESULTS

3.1. Threats and cooperation

3.1.1. Assignment of threats

Threats are widely used. Indeed, a minority of players refrains from using threats: they are only 16.25% in the Threat treatment (91 observations out of 560) and 12.64% in the Second Order treatment (91 observation out of 720). Figure 1 displays the average number of threat points assigned for each possible contribution level between zero and 20 by treatment. Overall, people threaten less as contribution increases. On average 7.34 and 6.68 threat points are assigned for a contribution level equal to zero in the Threat and Second Order treatments, respectively. The corresponding numbers are 0.66 and 0.33 threat points for the highest possible contribution of 20.

Interestingly, Figure 1 also shows that participants still threaten to punish very high contributions. For example, they assign 3.77 and 2.34 threat points on average for a contribution of 19 in the Threat and Second Order treatments, respectively. The data shows that the threshold of contribution from which players cease threatening is also relatively high. 51.96% of the players in the Threat treatment and 38.47% in the Second Order treatment assign points up to a contribution level of 19. These findings suggest that people use threats to signal that the group members should coordinate on the highest possible contribution. But threats also reveal to some extent the existence of anti-social behavior. Indeed, in 11.61% of the observations in the Threat treatment and 6.53% in the Second Order treatment threat points are assigned for the highest possible contribution.

Turning next to the differences across treatments, Figure 1 indicates that for all possible contribution levels, the average threat is higher in the Threat treatment than in the Second Order treatment. Our findings regarding threat decisions are summarized in Result 1.

[Figure 1 and Table 2 about here]

RESULT 1: *Most people threaten. The severity of threats decreases in contributions but people threat up to a high level of contribution. For all contribution levels, people assign significantly less threat points in the Second Order treatment than in the Threat treatment. There is an escalation of threats over time, except in the last period.*

Support for Result 1: Complementing the descriptive statistics reported above, Table 2 contains the estimates of various regression models. Model (1) is a random-effects Probit model in which the dependent variable is the probability to threaten group members. Random-effects models are justified since the same subjects play repeatedly. Model (2) is a random-effects Generalized Least Square model that estimates the determinants of the threshold of contribution from which the player no longer threatens to sanction. Models (3) to (7) are random-effects GLS models with robust standard errors and clustering at the individual level in which the dependent variable is the number of threat points that a player assigns for a given level of contribution *c*. *c* takes the following values in Table 1: $c = \overline{c}$, c = 0, 10, 15, and 20. Standard errors are clustered at the individual level to correct for the correlation of residuals across observations and for heteroskedasticity. In all of the regressions, the independent variables include the Second

Order treatment with the Threat treatment as the reference category, a time trend and a dummy variable for the final period.

The results of the first regression indicate that people are as likely to threaten others in the Threat and in the Second Order treatment and this likelihood does not change over time. Similarly, the threshold from which people cease threatening others does not differ across treatments and does not evolve over time. In contrast, people assign significantly less threat points in the Second Order treatment than in the Threat treatment for any positive contribution level (but not for $c = 0$). A possible reason behind the assignment of less threat points in the Second Order treatment is that subjects may anticipate that any difference between announcements and actual punishment will be publicly observed and sanctioned in the last stage of the game.

Table 2 also indicates an escalation of threats over time as the time trend is significant for all contribution levels except for the highest one. An interpretation is that threats become less and less effective and that people tend to compensate by increasing their severity, except in the final period. But do threats ever influence contribution decisions?

3.1.2 Contributions

Figure 2 displays the time path of individual contributions by period, averaged across groups, in the different treatments. Our observations regarding contribution levels are described as Result 2.

[Figure 2 about here]

RESULT 2: *In the Threat treatment, non-binding threats of punishment increase the average contribution level compared with the Baseline treatment; there is no evidence of crowding-out of the motivation to cooperate. However, the possibility of observing individual threat and sanction patterns and the introduction of a second round of punishment hurts cooperation.*

Support for Result 2: As shown by Figure 2, introducing non-binding threats of punishment has a positive effect on contribution levels in all periods. The average contribution levels are highest in the Threat treatment (mean = 18.19 ECU per individual from a maximum possible of 20, S.D. = 3.315), followed by the Baseline (16.05 ECU, S.D. = 5.00), and by the second order treatment (15.95 ECU, S.D. = 4.90). Two-tailed Mann-Whitney pairwise tests, with each group average contribution over the session as an independent observation, indicate that the difference in contributions between the Baseline and Threat treatments is significant ($p = 0.06$) as well as the difference between the Threat and the Second Order treatments ($p = 0.08$). In contrast, there is no significant difference between the Baseline and the Second Order treatments ($p > 0.010$).

To identify the determinants of contributions, we have estimated several regressions in which the dependent variable is the player's contribution. Table 3 reports the results of these estimations. Regressions (1) and (2) have been estimated by means of a random-effects Generalized Least Squared model, with robust standard errors and clustering at the individual level. In regressions (3) to (5) we use instead random-effects Tobit specifications to check the robustness of our results and to account for both the left- and right-censoring of observations. Last, regression (6) reports the estimation of a Tobit model on the period 1 data pooled across treatments to capture the pure effect of threat announcements on cooperation. The independent variables include several dummy variables to control for treatment effects, a time trend and a dummy variable for the final period. When the data from all the treatments are pooled together (regressions 1, 3, and 6), the omitted variable is the Baseline treatment. The independent variables also include the total number of threat points received from the three other group members averaged on all possible contribution levels and the total number of threat points received for the highest possible contribution of 20, respectively. They also include the threshold from which the player no longer assigns threat points to his group members and a dummy variable indicating whether the player threatens others for the highest possible contribution.

[Table 3 about here]

The positive and significant coefficients associated with the Threat variable in estimates (1) and (3) indicate that the participants contribute more in the Threat treatment than in the Baseline. On average individuals invest 2.141 ECU more in the group account in the Threat treatment (regression (1)). Interestingly, participating in the Threat treatment makes a significant positive difference on contributions from the very beginning of the game, as indicated by model (6). In contrast, controlling for the intensity of threats received, players contribute significantly less (-1.908 ECU) in the Second Order treatment than in the Threat treatment (regression (2)). The estimation of the various Tobit models confirms these findings.

Models (2) and (4) also show that the observation of the average threats announced by the other group members affects the level of contribution significantly and positively. In contrast, controlling for the general impact of threats, model (5) reveals that players react to anti-social threats (those directed towards the highest possible contribution of 20 ECU) by reducing their contribution significantly. We also find that the higher the threshold from which subjects no longer assign threat points, the more they cooperate. In contrast, those who assign threat points to the highest possible contribution contribute significantly less, which gives some support to the notion of anti-social threats. Last, contribution levels increase significantly over time, as reported in several previous studies on VCM games with sanctions; this is however significant in the Tobit regressions only. In all of the regressions, contributions decline in the final period of the game.

We did not include in these regressions the number of actual sanctions received in the previous period to avoid any autocorrelation problem. To measure their impact, we have estimated in separate random-effects GLS regressions (not reported here but available upon request) the determinants of changes in individual contributions between period *t* and period *t*+1. We used separate estimates for low contributors (those who contribute less than the group average in period t), and high contributors (who contribute more than the group average in period *t*) (N = 457 and 1291, resp.; $R^2 = 0.429$ and 0.081, resp.). We also included interaction variables between received points of sanctions and treatments, and the deviation between *i*'s and the others' average contributions. The estimates show that while sanctions raise contributions for individuals who contributed below the average (coeff. = 0.316, $p = 0.001$), they have no significant impact on those who contributed more than the average ($p = 0.635$). Punishment points do not have a different impact in the Threat and the Second Order treatments than in the Baseline ($p = 0.763$ and $p = 0.487$ for the low contributors, $p = 0.881$ and $p = 0.374$ for the high contributors, resp.). In similar regressions on the sole Second Order treatment, we have also investigated the effect of the sanctions received in the second punishment stage. The results indicate no significant effect of this additional variable (low contributors: $p = 0.178$, N = 198, $R^2 = 0.546$; high contributors: $p = 0.191$, N=284, $R^2 = 0.105$), suggesting that receiving sanctions in the last stage of this treatment is not interpreted as punishment for low contribution.

3.2. Threats and sanctions

3.2.1. Threats and first order punishment

Are threat decisions a strong predictor of future punishment decisions? Figure 3 displays the evolution over time of both the average number of threat points assigned and the average number of punishment points actually assigned in the first round of sanctions by treatment.

[Figure 3 about here]

Figure 3 indicates that subjects use costly punishment in all treatments and that punishment declines over time. It also shows that actual sanctions are less severe as announced. Our findings are stated more precisely in Result 3.

RESULT 3. *In all treatments, people assign costly punishment points but the assignment of sanctions declines over time whereas the assignment of threats increases over time. Although people contribute more in the Threat treatment than in the other treatments, punishment is weakly higher in this treatment as if people feel committed to their announcements. Finally our data indicate that while sanctions are less severe than announced, threats are nevertheless strong predictors of subsequent sanctions.*

Support for Result 3: On average the subjects actually assign on average 0.423 punishment point in the Baseline treatment (S.D. = 1.423), 0.607 point in the Threat treatment (S.D. = 1.761), and 0.454 in the Second Order treatment (S.D. = 1.479). Mann-Whitney pairwise tests, with each group decision as an observation, conclude that there is no difference in punishment levels between the Threat and the Baseline treatments (z =-0.38, $p > 0.100$), between the Second Order and the Baseline treatments (z =-0.795, $p > 0.100$), or between the Second Order and the Threat treatments (z = 0.476, $p > 0.100$). These tests do not control, however, for the amounts contributed.

The left panel of Table 4 complements these findings by reporting the estimates of two random-effects Tobit models in which the dependent variable is the number of punishment points that player *i* assigns to player *j* in the (first) punishment stage of period *t*. The first model pools the data of the three treatments, while the second model pools the data of the Threat and Second Order treatments only. The independent variables include dummy variables for each treatment, the average amount contributed by the group (excluding *j*'s contribution), the differences between *j*'s and the group average contribution, conditional on *j* contributing less or more than the group average, a time trend, and a dummy variable for the final period. In the second model, they also include the amount of threats assigned by *i* for the amount of contribution corresponding to *j*'s actual contribution to measure whether individuals tend to respect their threats when they actually punish. In addition, a variable indicates whether the subject *i* has sent threat points for the highest possible contribution as an index of anti-social behavior.

[Table 4 about here]

As demonstrated in previous studies, Table 4 indicates that players receive more punishment points, the less they have contributed relative to their group average. Model (1) shows that, controlling for differences between the target's and the average contribution in the group, players punish marginally more in the Threat treatment than in the Baseline treatment. How can one explain that subjects punish more in the Threat treatment whereas Result 1 has shown that players cooperate more in this treatment? Indeed, higher contributions should lead to less sanctions. This effect is partly offset by the fact that the players are incited to punish more in order to fulfill their (yet non-binding) announcements. This is confirmed by the estimation of model (2) indicating that the more threat points announced, the more punishment points actually assigned. This indicates that threats should be eventually considered as credible signals of subsequent sanctioning decisions. Model (2) also indicates that the subjects who threaten to punish the highest contribution level are also more willing to sanction others.

3.2.2. Threats and second order punishment

Previous findings have shown that threats have a positive effect on cooperation. However, allowing people to observe the individual difference between the announcement of threats and the sanctions actually assigned and introducing an additional round of punishment seem to destroy this effect. In this section we investigate the determinants of the subjects' decisions to sanction in the second round of punishment and the incidence of second order punishment on further threats.

In the last stage of the Second Order treatment, subjects may indeed sanction the individual differences between the announced threats and the actual sanctions assigned by their group members. But they may be willing to sanction them for several other possible reasons. They may sanction second order free riders (i.e. those who failed to sanction low contributors in stage three whatever their threat announcements). They may also punish a player who punishes more or less than the average, or even a player who has punished group members who contribute more than the average (perverse punishers). They may counterpunish for having received punishment points in the first round of sanction. Note that revenge can only be blind since individuals are never informed about who threatened and sanctioned them personally. However subjects may use information on the severity of sanctions directed toward each player other than herself as a signal of whom could have punished her. Our results regarding the determinants of second round punishment are summarized below.

RESULT 4. *After controlling for several possible motives for sanctioning in the last round of punishment in the Second Order treatment, our data shows that people use the second round of punishment to sanction those who assign less points than announced in the first stage of the game.*

Support for Result 4. We consider the influence of each of the possible determinants described above in the three regressions reported in the right panel of Table 4. The models estimated are random-effects Tobit models accounting for the left-censoring of the data. The dependent variable is the number of punishment points that player *i* assigns to player *j* in the second round of sanctions in period *t*. In column (3), the independent variables include the average group contribution (excluding *j*'s contribution) and the absolute values of positive and of negative differences between *j*'s contribution and the average contribution of the other group members if *j* contributes more or less, respectively, than the average of the group (these variables are equal to 0 otherwise). They also include the average number of threat points assigned by player *j* to others (except player *i*) corresponding to their actual contribution levels, and the number of punishment points actually assigned by *j* to these group members (except player *i*). A dummy variable captures the impact of player *j* punishing less than the announced threats. To identify whether blind revenge could be at play, a dummy variable indicates whether player *i* has been punished or not in the first round of sanctions.

In model (4) we add two variables to capture sanction enforcement. More precisely, the first one takes the value of the positive difference between the average number of punishment points assigned by player *j* to his group members (excluding *i*) when *j* punishes more than the average of the group (excluding *i*), and 0 otherwise. More formally, this writes: $\max\left\{\sum_{k\neq i} p_j^{k1t} - \left(\sum_{m\neq j}\sum_{k\neq i,j} p_m^{k1t}\right)/2, 0\right\}$. The second additional variable takes the absolute value of the negative difference between the average number of punishment points assigned by player *j* to his group members (excluding *i*) when *j* punishes less than the average of the group (excluding *i*), and 0 otherwise. This writes $\max\left\{0, \left(\sum_{m\neq j}\sum_{k\neq i,j} p_m^{k1t}\right)/2 - \sum_{k\neq i} p_j^{k1t}\right\}$. These variables should identify the punishment of second order free riding. Model (5) is equivalent to model (4) except that a dummy variable indicates whether player *i* has threatened his group members for the highest possible contribution, to measure whether anti-social threatening behavior is associated with a specific punishment behavior in the final stage of the game.

Overall, and in contrast with the first-round punishment, second-round punishment is more likely to occur when the group has established a norm of cooperation since the coefficient of the average group contribution variable is positive. The three estimations confirm that a subject is more likely to be punished in the second round of sanctions when he has actually assigned less punishment points than threat points. But we also find some evidence of other motives to punish in the final stage. In particular, low contributions are still punished in the second round of sanctions as indicated by the significant coefficient associated with the negative difference between *j*'s contribution and the group contribution. We also find some evidence of blind revenge as the subjects who have been punished in the first round of sanctions are more likely to counter-punish in the second round of sanction, although the target in the second round may not be the subject's punisher in the first round. Additional support to blind revenge can be found in the significant coefficient of the variable indicating the number of punishment points assigned by player *j*. Indeed, the assignment by player *j* of many punishment points may be used as a signal that *j* is at the origin of *i*'s being punished. In contrast, the variables associated with sanction enforcement turn out to be insignificant. Last, we find that the "anti-social threateners" are also more likely to punish in the second round of sanctions.[5]

3.2.3. Implications of second-round punishment on further threats

Since deviations between threats and actual sanctions are punished, subjects in this treatment are expected to react by adjusting either their further punishment behavior

[5] In an additional regression (not reported here but available upon request), we have tested whether players punish perverse first-order punishers (i.e. those who have sanctioned group members who contributed more than the average). The coefficient of this variable is however not significant, indicating that second order punishment is not used to deter perverse punishment, in contrast with Cinyabuguma (2006).

upward or their threat pattern downward. Since second-round punishment hurts more heavily those who punish more (as indicated by Table 4) and since punishment is costly, it is more likely that subjects adjust their threat downward. Our findings are summarized in Result 5.

RESULT 5. *In the Second Order treatment, subjects who threaten more than they actually punish and who are punished in the last stage of period t revise their threats downward in the next period.*

Support for Result 5. We have estimated the determinants of changes in the total number of threat points assigned by a subject to his group members between periods *t* and *t+1* by means of a random-effects GLS model with robust standard errors and clustering at the individual level (not reported here, but available upon request). This model is estimated separately for the subjects who threatened more than they actually punished in period *t* (N = 711, R^2 = 0.120) and for those who actually assigned more or the same number of punishment points than threat points (N = 1341, R^2 = 0.002). The independent variables consist of both the difference between the number of threat points and the actual sanctions assigned by player *i* to his group members after being informed on their contribution levels, and the total number of punishment points received by player *i* in stage four of period *t*.

We find that those individuals who distributed more threat points than punishment points in stage three of period *t* respond to sanctions received in stage four by revising downward the number of threat points they assign in the following period (coeff. = -0.170, *p* = 0.028). Moreover, the more they deviated in period *t*, the more they revise downward (coeff. = -0.452, $p < 0.001$). No such adjustment is observed for those who punished either according to their threats or more severely than their threats ($p = 0.570$ and $p = 0.814$, respectively). This tendency to reduce the deviations between announcements in stage one and actual sanctions in stage three by revising threats downward could explain that in this treatment threats do not improve cooperation in comparison with the Threat treatment.

3.4. Efficiency

In this section we investigate the consequences of threats and second order punishment on efficiency. If threats are sufficient to induce higher cooperation, then sanctions need not be implemented, which should reduce the detrimental effects of punishment on efficiency and improve welfare. However the data contradicts this hypothesis. This is summarized in Result 6.

RESULT 6. *While threats do improve cooperation in the Threat treatment, they do not increase efficiency before the second half of the game. The possibility to observe deviations between threats and sanctions and to assign a second round of sanctions in the Second Order treatment decreases efficiency.*

Support for Result 6. Comparing the before-sanction payoffs in the Baseline and the other treatments indicates that threats induce a positive effect on welfare if the individual deviations between threats and actual sanctions cannot be observed. Indeed, the mean payoffs amount to 29.63 ECU in the Baseline treatment (S.D. = 4.95), 30.92 in the Threat treatment (S.D. = 3.45), and 29.57 ECU in the Second Order treatment (S.D. = 5.20). However, the positive effect of threats on cooperation is offset by the cost of sanctions.

The direct cost of punishment can be easily measured by comparing the before-sanctions and after-sanctions payoffs in each treatment. The final payoffs amount to 25.84 ECU in the Baseline treatment (S.D. = 8.31; this corresponds to 87.21% of the before-sanctions payoff), 25.47 ECU in the Threat treatment (S.D. = 9.31; 82.37% of the before-sanctions payoff), and 23.20 ECU in the Second Order treatment (S.D. = 11.17; 78.46% of the before-sanctions payoff). Subjects in the Threat treatment try to fulfill their commitment by assigning more punishment points compared with the Baseline treatment. As a consequence, more punishment induces more cooperation, by inciting the free riders to contribute more, but also impose higher social costs. The relative loss induced by the Second Order treatment results both from a lower incentive effect of threats on contributions and from higher costs of punishment due to the existence of an additional stage of sanction.

A formal proof of these results is given in Table 5. Table 5 reports the estimations of three GLS models on pooled data with robust standard errors and clustering at the individual level in which the dependent variable is the before-sanction payoff (model (1)) or the after-sanction payoff (models (2) and (3)). The independent variables include each treatment, with the Baseline as the omitted reference category, a time trend and a dummy variable for the last period.

[Table 5 about here]

These regressions indicate that the Threat treatment induces significantly higher before-sanction payoffs than the Baseline treatments (model (1)). A positive effect on welfare is also observed through after-sanction payoffs but only in the second half of the game (model (3)). Finally Table 5 indicates that in the Second Order treatment efficiency does not differ from the Baseline treatment if we consider the before-sanction payoffs but it is significantly lower if one considers the after-sanction payoffs.

3.4. A robustness check

To test the robustness of our results to changes in the parameters of the game, the same three treatments have been run in a low cost condition in which one punishment point costs one ECU to both the punisher and the target instead of two for the target. How are threats, cooperation and efficiency affected by the change in the cost of sanctions to punishees? Consistent with our previous results, we find that in the low condition also the average individual contributions are the highest in the Threat treatment (11.51, S.D. = 2.13), followed by the Baseline treatment (10.07, S.D. = 6.08), and by the Second Order treatment (8.86, S.D. = 5.39). Figure 4 displays the evolution of individual contributions over time by treatment in the low-cost condition. It shows that the effect of threats on average contributions is less persistent over time than in the high cost condition. Our findings regarding the low cost condition are summarized in Result 7.

[*Figure 4 about here*]

RESULT 7: *The number of threat points assigned to group members for almost any contribution level does not differ in the low-cost and the high-cost conditions of each treatment. In both conditions, the threat of punishment in the Threat treatment has a positive effect on the average contributions compared with the Baseline treatment. This effect is, however, less persistent over time in the low-cost than in the high-cost condition.*

Overall, efficiency is not increased by the introduction of threats in this condition and earnings are even lower in the low-cost than in the high-cost condition.

Support for Result 7: GLS estimations with robust standard errors and clustering at the individual level (not reported here but available upon request) indicate that the threshold of contribution from which a subject no longer assigns threat points is similar in the low-cost and high-cost conditions of the Threat treatment (N = 1280; *p* = 0.651) and of the Second Order treatment (N = 1440; *p* = 0.274). The same conclusion is reached for every level of contribution in both treatments (*p* > 0.100), except that threats against the maximum contribution are higher in the low-cost condition than in the high-cost condition of the Second Order treatment (N =1440; *p* = 0.037). In other words, in this treatment anti-social threatening behavior is more frequent in the low-cost condition possibly because second order punishment is expected to be less likely in this condition.

As regards contributions, a Mann-Whitney pairwise test comparing average contributions in the Threat and the Baseline treatments in the low cost condition indicates that people contribute significantly more in the Threat treatment than in the Baseline in the first ten periods only (*p* = 0.070). No significant difference is found between these treatments after period 10. Similarly, the effect of threats in the Second Order treatment is weaker and less persistent than in the high cost condition. Indeed while the average contribution is higher in the Threat treatment than the Second Order treatment in the first ten periods (*p* = 0.050), no significant difference is found in the second half of the game.

Regarding differences across conditions, a Mann-Whitney test comparing contributions in the Baseline treatment in the high-cost (16.05 ECU) and the low-cost (10.07 ECU) conditions indicates that people contribute significantly more in the high-cost condition (p = 0.007). A similar test comparing the contributions in the Threat treatment in the low-cost condition (11.51 ECU) and the high-cost condition (18.19 ECU) reach the same conclusion (p = 0.053). Similar results are obtained when comparing contributions in the Second Order treatment in the low-cost (8.86 ECU) and high-cost condition (15.95 ECU) (p = 0.012). While the level of threats is similar across conditions in each treatment, and as reported in previous studies, cooperation is reduced when the payoff reduction of punishment is lower.

Mann-Whitney pairwise tests indicate that average earnings before the punishment stage are not higher in the Threat treatment than in the Baseline treatment (26.91 and 26.04 respectively, p = 0.453) if all periods are considered together. They are also similar in the Second Order treatment (25.32) than in both the Baseline (p = 0.627) and the Threat treatment (p = 0.233). If the mean final payoffs are considered instead, there is no difference between the Baseline treatment (22.42) and the Threat treatment (22.97, p = 0.965), while payoffs are significantly lower in the Second Order treatment than in both the Baseline (16.29; p = 0.015) and in the Threat treatment (p = 0.024).

Regarding differences across conditions, the average before-punishment earnings are smaller in the low-cost than in the high-cost condition in the Baseline (26.04 and 29.63 ECU, respectively, p = 0.007), the Threat treatment (26.91 and 30.92 ECU, p = 0.003), and the Second Order treatment (25.32 and 29.57 ECU, p = 0.012). This is due to the fact that although people receive the same quantity of threats, they contribute less. The same conclusions are reached if one considers instead the final earnings after the punishment stage in the Baseline treatment (22.42 and 25.84 ECU; *p* = 0.101) and the Second Order treatment (16.29 and 23.20 ECU; *p* = 0.038). The comparison of final earnings in the Threat treatment also indicates that earnings are smaller in the low-cost condition, but not significantly so (22.97 and 25.47 ECU; *p* = 0.315).

4. CONCLUSION

Usually threats preexist to the enforcement of punishment. However, most experiments on public good games with sanctions have so far ignored the potential impact of such threats. We have designed an experiment to analyze whether individuals are willing to threaten before sanctioning and to measure the influence of such threats on contributions and efficiency. While the Baseline treatment replicates the standard VCM game with sanctions, we introduce in the Threat treatment a preliminary stage in which subjects can assign non-binding threat points for each contribution level. The Second Order treatment adds to the Threat treatment a final stage in which subjects can observe and punish the differences between the threats and actual sanctions of each group member.

We find that most individuals threaten up to a high level of contribution, although less severely in the Second Order than in the Threat treatment. In the Threat treatment, these non-binding threats increase the average contribution level compared with the Baseline treatment, while the possibility of observing individual threats and actual sanctions and the introduction of a second round of punishment hurt cooperation. Although subjects cooperate more in the Threat treatment than in the other treatments, they also punish more as if people feel committed to their announcements. While sanctions are less severe than announced, threats are nevertheless strong predictors of subsequent sanctions. As a consequence, threats cannot increase efficiency before the second half of the game. Efficiency is even constantly lower in the Second Order treatment where people can observe the differences between individual threats and actual sanctions. In this treatment, people adjust their threats and actual sanctions to avoid second order punishment by reducing their threats. These results are relatively robust to a change in the monetary consequences of sanctions for the punished individuals. In the low-cost condition of the Threat treatment, the positive effect of threats on cooperation is, however, less persistent over time than in the high-cost condition.

Overall, if one compares these findings with the previous literature on pre-play communication, it seems that non-binding contributions are more efficient than non-binding threats of sanctions.

REFERENCES

Anderson, C.M. and L. Putterman. 2006. "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," *Games and Economic Behavior*, *54*(1), 1-24.

Andreoni, J. 1988. "Why Free Ride: Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 35 (1), 57-73.

Bochet, O., T. Page, and L. Putterman. 2006. "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization*, 60(1), 11-26.

Bochet, O., and L. Putterman. 2009. "Not just babble: Opening the black box of communication in a voluntary contribution experiment," *European Economic Review*, 53(3), 309-326.

Brosig, J., A. Ockenfels, and J. Weimann, 2003. "The effect of communication media on cooperation," *German Economic Review*, 4, 217–242.

Carpenter, J.P. 2007a. "Punishing Free Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods,*"* *Games and Economic Behavior*, 60(1), 31-51.

_______ 2007b. " The demand for punishment," *Journal of Economic Behavior and Organization* 62, 522–542.

Cinyabuguma, M., T. Page, and L. Putterman. 2006. "Can Second Order Punishment Deter Perverse Punishment?," *Experimental Economics*, 9(3), 265-279.

Denant-Boemont, L., D. Masclet and C. Noussair 2007. "Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment," *Economic Theory*, 31(1), 145-167.

Dickinson, D., and M.C. Villeval. 2008. "Does Monitoring Decrease Work Effort? The Complementarity Between Agency and Crowding-Out Theories," *Games and Economic Behavior*, 63 (1), 56-76.

Duffy, J., and N. Feltovich. 2006. "Words, deeds, and lies: strategic behaviour in games with multiple signals ," *The Review of Economic Studies* 73 (3), 669-688.

Egas, M. and A. Riedl. 2008. "The economics of altruistic punishment and the maintenance of cooperation," *Proceedings of the Royal Society B - Biological Sciences*, 275 (1637), 871-878.

Fehr E., and S. Gächter 2000. "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), 980-94.

Fehr, E., and S. Gächter. 2002. "Altruistic punishment in humans," *Nature*, 415, 10 January, 137-140.

Fehr, E., and B. Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism," *Nature*, *422*, 137-40.

Fischbacher, U. 2007. “Z-Tree: Zurich Toolbox for Ready-made Economic experiments,” *Experimental Economics*, 10(2), 171-178.

Houser, D., E. Xiao, K. McCabe, and V. Smith. 2008. "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation," *Games and Economic Behavior,* 62(2), 509-532.

_______. 2007. “Money, religion and revolution,” *Economics of Governance*, 8(1), 1-16.

Isaac, R. M., K. McCue, and C. Plott. 1985. “Public Goods Provision in an Experimental Environment,” *Journal of Public Economics*, 26 (1), 51–74.

Isaac, R. M., and J.M. Walker. 1988a. “Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism,” *Quarterly Journal of Economics,* 103 (1), 179-99.

_______. 1988b. “Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism,” *Economic Inquiry*, 26(4), 585-608.

_______. 1991. “Costly Communication: An Experiment in a Nested Public Goods Problem," in T. Palfrey (Ed.). *Contemporary Laboratory Research in Political Economy*. Ann Arbor: Univ. of Michigan Press.

Kerr, N.L., and C.M. Kaufman-Gilliland. 1994. “Communication, commitment, and cooperation in social dilemmas,” *Journal of Personality and Social Psychology*, 66, 513-529.

Krishnamurthy, S. 2001. “Communication Effects In Public Good Games With And Without Provision Points,” in M. Isaac (Ed.). *Research In Experimental Economics*, Volume Eight, Amsterdam : JAI.

Ledyard J. 1995. “Public Goods: A Survey of Experimental Research”, in Kagel J. and Roth. A., Eds., *Handbook of Experimental Economics*. Princeton, Princeton University Press, 111-194.

Li, J., E. Xiao, D. Houser, and P.R. Montague. 2009. “Neural responses to sanction threats in two-party economic exchange,” *PNAS*, 106(39), 29 September, 16835-16840.

Marwell, G., and R.E. Ames. 1979. "Experiments on the provision of public goods. I: Resources, interest, group size, and the free-rider problem," *American Journal of Sociology,* 84(6), 1335–1360.

Masclet, D., C. Noussair, S. Tucker and M.C Villeval. 2003. “Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism,” *American Economic Review*, 93 (1), 366-380.

Nikiforakis, N.S. 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?," *Journal of Public Economics*, 92, 91–112.

Noussair, C., and S. Tucker. 2005. “Combining Monetary and Social Sanctions to Promote Cooperation,” *Economic Inquiry*, 43 (3), 649-660.

Ostrom, E., J. Walker, and R. Gardner. 1992. "Covenants With and Without a Sword: Self-Governance Is Possible," *American Political Science Review*, 86(2), 404–417.

Sally, D. 1995. Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to1992. *Rationality and Society*, 7, 58-92.

Sefton, M., R. Shupp, and J. Walker, 2007. “The Effect of Rewards and Sanctions in Provision of Public Goods,” *Economic Inquiry*, Vol. 45, pp. 671-690.

Yamagishi, T. 1986. “The Provision of a Sanctioning System as a Public Good,” *Journal of Personality and Social Psychology* 51(1) pp. 110-16.

Table 1. Characteristics of the experimental sessions

Session number	*# subjects*	*# groups*	*Treatment*	*Condition*
1	12	3	Baseline	High-cost
2	16	4	Baseline	High-cost
3	20	5	Threat	High-cost
4	8	2	Threat	High-cost
5	12	3	SdOrder	High-cost
6	12	3	SdOrder	High-cost
7	12	3	SdOrder	High-cost
8	12	3	Baseline	Low-cost
9	12	3	Baseline	Low-cost
10	12	3	Baseline	Low-cost
11	12	3	Threat	Low-cost
12	12	3	Threat	Low-cost
13	12	3	Threat	Low-cost
14	12	3	SdOrder	Low-cost
15	12	3	SdOrder	Low-cost
16	12	3	SdOrder	Low-cost
Total	200	50		

Table 2. Determinants of threats in the high cost condition

Dependent variables	Decision to threaten	Threshold of threats	Average number of threat points assigned				
			For any *c*	For *c*=0	For *c*=10	For *c*=15	For *c*=20
Models	RE Probit	RE GLS[a]	RE GLS[a]	RE GLS[a]	RE GLS[a]	RE GLS[a]	RE GLS[a]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Threat treatment	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Second Order treatment	0.041	-0.707	-3.038**	-1.977	-3.198**	-3.555**	-1.076*
	(0.533)	(1.448)	(1.215)	(1.404)	(1.391)	(1.429)	(0.584)
Period	-0.008	0.090	0.366***	0.193***	0.340***	0.556***	0.005
	(0.013)	(0.059)	(0.065)	(0.065)	(0.077)	(0.075)	(0.031)
Final period	-0.701**	-1.784***	-2.491***	-1.904***	-2.564***	-3.634***	-0.197
	(0.316)	(0.559)	(0.520)	(0.666)	(0.535)	(0.598)	(0.589)
Constant	2.811***	15.331***	13.624***	20.107***	15.468***	9.517***	1.965***
	(0.440)	(1.312)	(1.121)	(1.286)	(1.275)	(1.268)	(0.456)
# Obs.	1280	1280	1280	1280	1280	1280	1280
Log-likelihood	-270.439						
R^2		0.008	0.135	0.038	0.093	0.175	0.024
Rho	0.865	0.590	0.557	0.567	0.513	0.552	0.346

Notes: [a] RE GLS=Random Effects Generalized Least Squares. *** Significant at the 0.01 level; ** at the 0.05 level; * at the 0.1 level. Robust standard errors (in parentheses) are clustered at the individual level.

Table 3. Determinants of contribution in the high cost condition

Models	*RE GLS*[a]	*RE GLS*[a]	*RE Tobit*[b]	*RE Tobit*[b]	*RE Tobit*[b]	*Tobit*
Treatments	*All*	*All except Baseline*	*All*	*All except Baseline*	*All except Baseline*	*All Period 1*
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	*Ref.*	-	*Ref.*	-	-	*Ref.*
Threat treatment	2.141***	*Ref.*	8.639***	*Ref.*	*Ref.*	5.651***
	(0.817)		(2.643)			(1.975)
Second Order treatment	-0.098	-1.908**	0.275	-7.673***	-8.095***	2.742
	(1.027)	(0.829)	(2.427)	(2.519)	(2.382)	(1.975)
Average threat received	-	0.109***	-	0.306***	0.294***	-
		(0.036)		(0.079)	(0.082)	
Threat received for *c*=20	-	-	-	-	-0.255**	-
					(0.115)	
Threshold of threats assigned	-	-	-	-	0.191**	-
					(0.077)	
Threat assigned for *c*=20	-	-	-	-	-4.851***	-
					(1.585)	
Period	0.055	-0.030	0.353***	0.308***	0.290***	-
	(0.036)	(0.039)	(0.054)	(0.076)	(0.075)	-
Final period	-3.567***	-3.363***	-9.952***	-10.130***	-9.947***	-
	(0.750)	(0.953)	(1.389)	(1.742)	(1.732)	-
Constant	15.665***	16.158***	17.948***	22.121***	20.459***	12.893***
	(0.595)	(0.645)	(1.890)	(2.241)	(2.293)	(1.360)
Observations	1840	1280	1840	1280	1280	92
ρ	0.392	0.389	0.478	0.476	0.447	
Lef censored obs.			124	82	82	
Right censored obs.			1073	798	798	
Log likelihood			-3032.871	-1910.201	-1901.063	-233.961
R^2	0.044	0.100				

Notes: [a] Random-effects Generalized Least Squares; [b] random-effects Tobit; *** significant at the 0.01 level; ** at the 0.05 level; * at the 0.1 level; robust standard errors clustered at the individual level in parentheses.

Table 4. Determinants of the number of punishment points assigned by player *i* to player *j* in the first round and the second round of punishment in the high cost condition (random-effects Tobit estimates)

	First round of punishment		Second round of punishment		
Treatments	All	All except	Second Order treatment		
	treatments	Baseline			
	(1)	(2)	(3)	(4)	(5)
Baseline treatment	*Ref.*		-	-	-
Threat treatment	1.212*	*Ref.*	-	-	-
	(0.670)				
Sd Order treatment	0.383	-0.893	-	-	-
	(0.624)	(0.582)			
Average contribution	-0.221***	-0.160***	0.112**	0.112**	0.117**
of others	(0.033)	(0.044)	(0.055)	(0.057)	(0.056)
Absolute positive	-0.297***	-0.236***	0.001	<0.001	-0.003
diff. from average	(0.051)	(0.063)	(0.069)	(0.069)	(0.069)
Absolute negative	0.525***	0.407***	0.283***	0.283***	0.282***
diff. from average	(0.021)	(0.027)	(0.028)	(0.028)	(0.028)
Threat assigned	-	0.377***			
to j		(0.042)			
Anti-social	-	1.364***			0.906*
threatener		(0.380)			(0.490)
j's average	-	-	-0.033	-0.033	-0.028
threat			(0.081)	(0.081)	(0.081)
j's average	-	-	0.502***	0.498**	0.502**
punish. in first round			(0.098)	(0.227)	(0.225)
j threats more	-	-	0.777*	0.769*	0.702*
than he punishes			(0.425)	(0.428)	(0.427)
Received sanctions	-	-	1.390***	1.393***	1.356***
in first round			(0.345)	(0.349)	(0.346)
Pos. dev. of *j* from average	-	-		0.012	0.007
punishment in first round				(0.250)	(0.249)
Neg. dev. of *j* from average	-	-		0.023	0.028
punishment in first round				(0.127)	(0.127)
Period	-0.314***	-0.329***	-0.171***	-0.171***	-0.160***
	(0.019)	(0.023)	(0.029)	(0.030)	(0.030)
Final period	0.100	-0.455	-0.072	-0.073	-0.092
	(0.567)	(0.695)	(0.933)	(0.933)	(0.924)
Constant	-0.051***	-0.224	-6.866***	-6.875***	-7.024***
	(0.728)	(0.923)	(1.119)	(1.150)	(1.139)
# observations	5520	3840	2160	2160	2160
# left cens.obs.	4676	565	1892	1892	1892
# right cens.obs.	46	30	-	-	-
Log-likelihood	- 3212.445	-2204.581	-1069.927	-1069.911	-1068.230
ρ	0.521	0.486	0.375	0.375	0.355

Note: *** Significant at the 0.01 level; ** at the 0.05 level; * at the 0.1 level. Standard errors in parentheses.

Table 5. Determinants of payoffs (GLS models)

Dependent variable	Before-sanction payoffs (1)	After-sanction payoffs (2)	After-sanction payoffs (3)
Baseline treatment	*Ref.*	*Ref.*	*Ref.*
Threat treatment	1.318***	-0.310	-0.951
	(0.355)	(0.927)	(1.026)
Threat*last 10 periods			1.282***
			(0.389)
Second Order treatment	-0.059	-2.639***	-2.639***
	(0.419)	(0.937)	(0.947)
Period	0.033**	0.526***	0.495***
	(0.015)	(0.041)	(0.020)
Final period	-2.140***	-4.678***	-4.575***
	(0.370)	(0.425)	(0.468)
Constant	29.399***	20.555***	20.873***
	(0.294)	(0.843)	(0.739)
# of observations	5520	5520	5520
R^2	0.025	0.096	

Note: *** significant at the 0.01 level; ** at the 0.05 level; * at the 0.1 level. Robust standard errors in parentheses with clustering at the individual level.

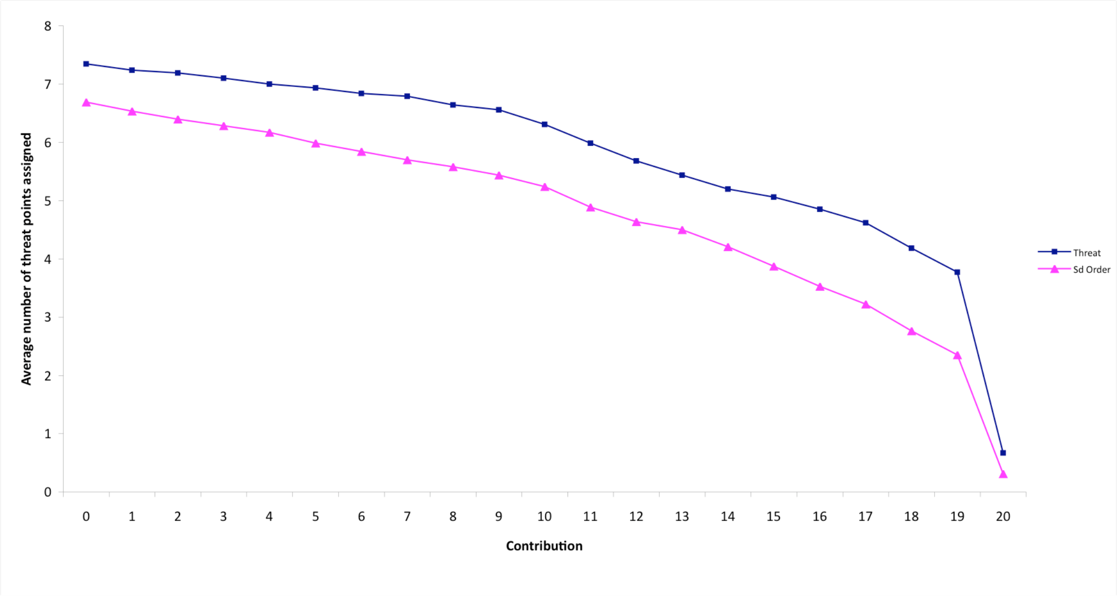


Figure 1. Average number of threat points assigned for each contribution level by treatment in the high cost condition

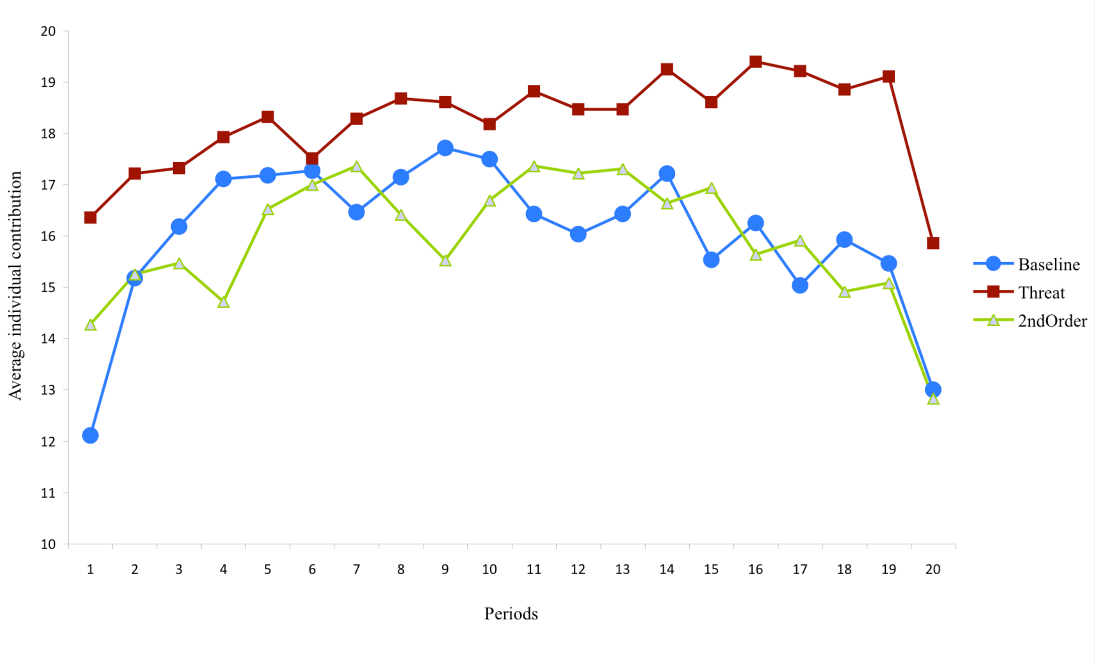


Figure 2. Evolution of the average individual contributions over time by treatment in the high cost condition

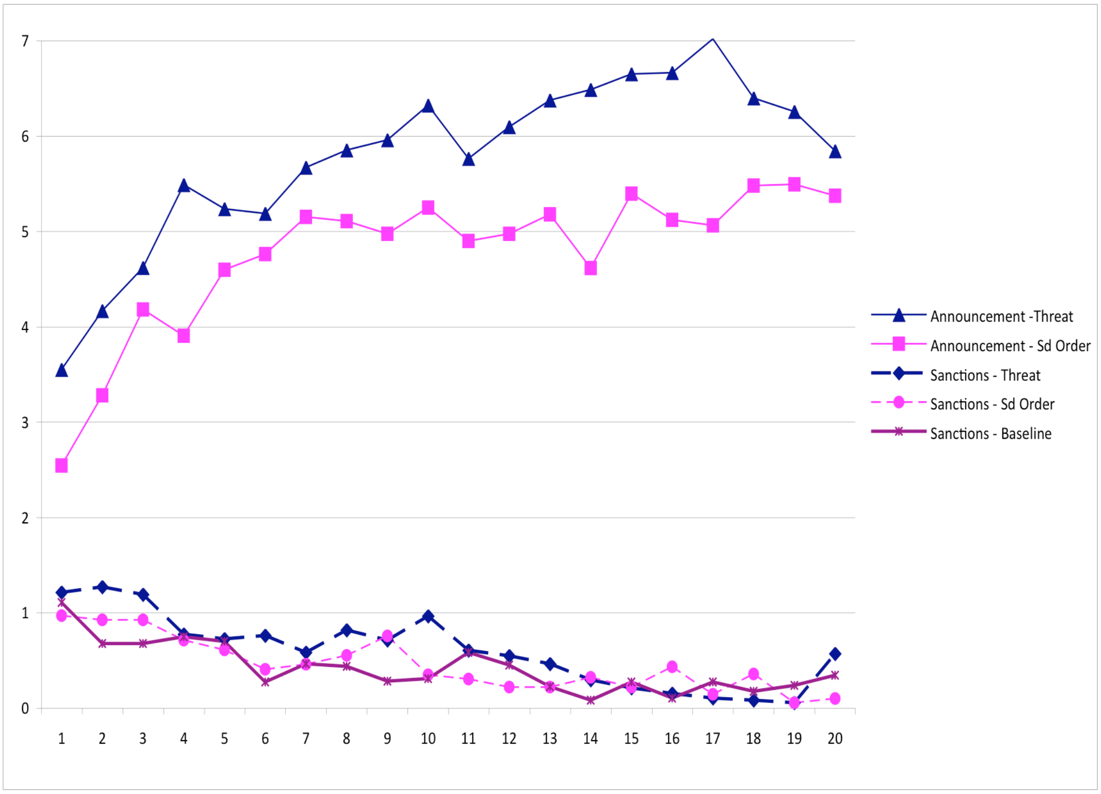


Figure 3. Evolution of threats and actual punishment over time by treatment

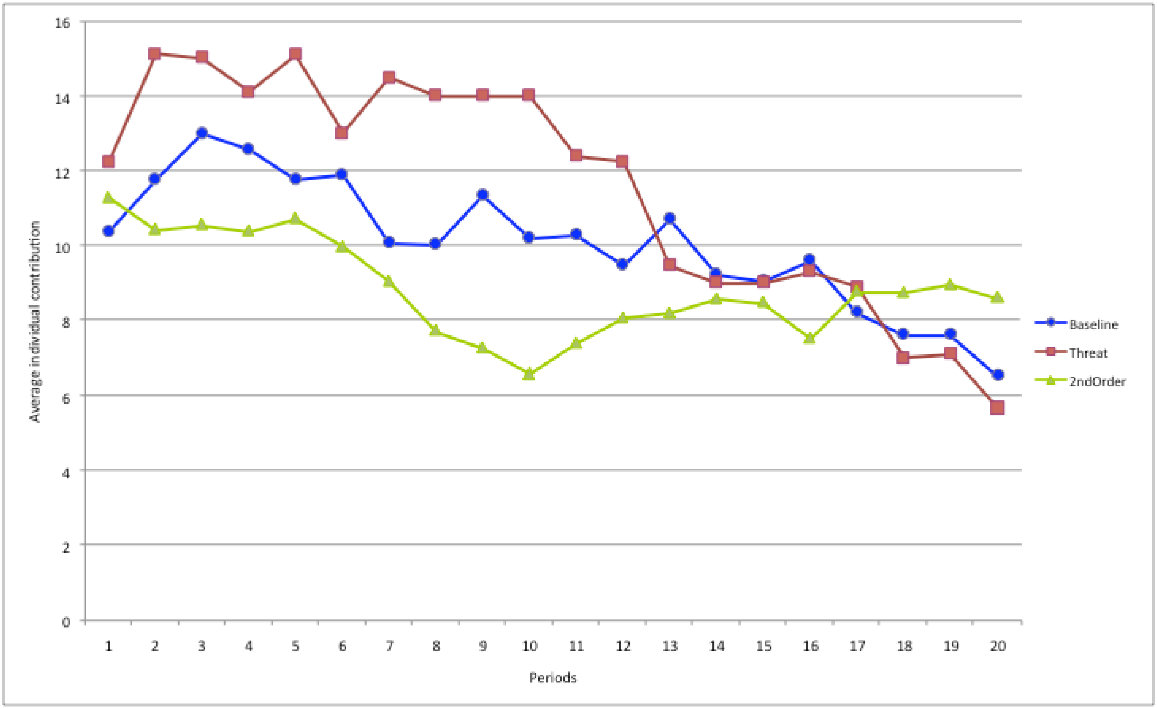


Figure 4. Evolution of the average individual contributions over time by treatment in the low cost condition

Appendix. Instructions of the Threat treatment (high-cost condition) *(The instructions for the other treatments are available upon request)*

You are taking part in an experiment in economics during which you can earn money. Your earnings depend on your decisions and on the decisions of the other participants with whom you will interact. It is therefore important to read these instructions with attention.

All the transactions during the experiment and your entire earnings will be calculated in ECU (Experimental Currency Units). At the end of the experiment the total amount of ECU you have earned during this session will be converted to Euros and paid to you in cash in a separate room by somebody who is not aware of the content of the experiment, according to the following rules:

- ❑ Your final payoff in ECU consists of the sum of your payoffs in each of the 20 periods comprising this session.
- ❑ This final payoff in ECU will be converted into Euros at the rate: 100 ECU = 2 Euros.
- ❑ In addition, you will be given a show up fee of 5 Euros.

At the beginning of the session, the participants are divided into groups of four. You will therefore interact with three other participants. **During the 20 periods, you will interact with the same persons**. You will never be informed of the identity of these persons.

Description of each period

In each period, after receiving an endowment of 20 ECU each, the four participants belonging to a group can participate in a project, by contributing to a group account that will be shared among them. The amount of this group account is determined by the sum of the individual contributions of the four members of the group. Next, the group members can indicate their disapproval to the contribution of other group members by assigning points that reduce their payoff. Each period consists of three stages:

- During the first stage, each group member indicates how many disapproval points he would be ready to assign to other group members for each possible contribution level in the second stage.
- During the second stage, after being informed on the number of disapproval points that the other group members propose to assign for each possible contribution level, each of the four group members decides simultaneously on his actual contribution to the project.
- During the third stage, after being informed on the individual contributions of the other group members, each one decides on the number of disapproval points he actually assigns to other group members and their payoffs are reduced accordingly.

The details of each stage are described below.

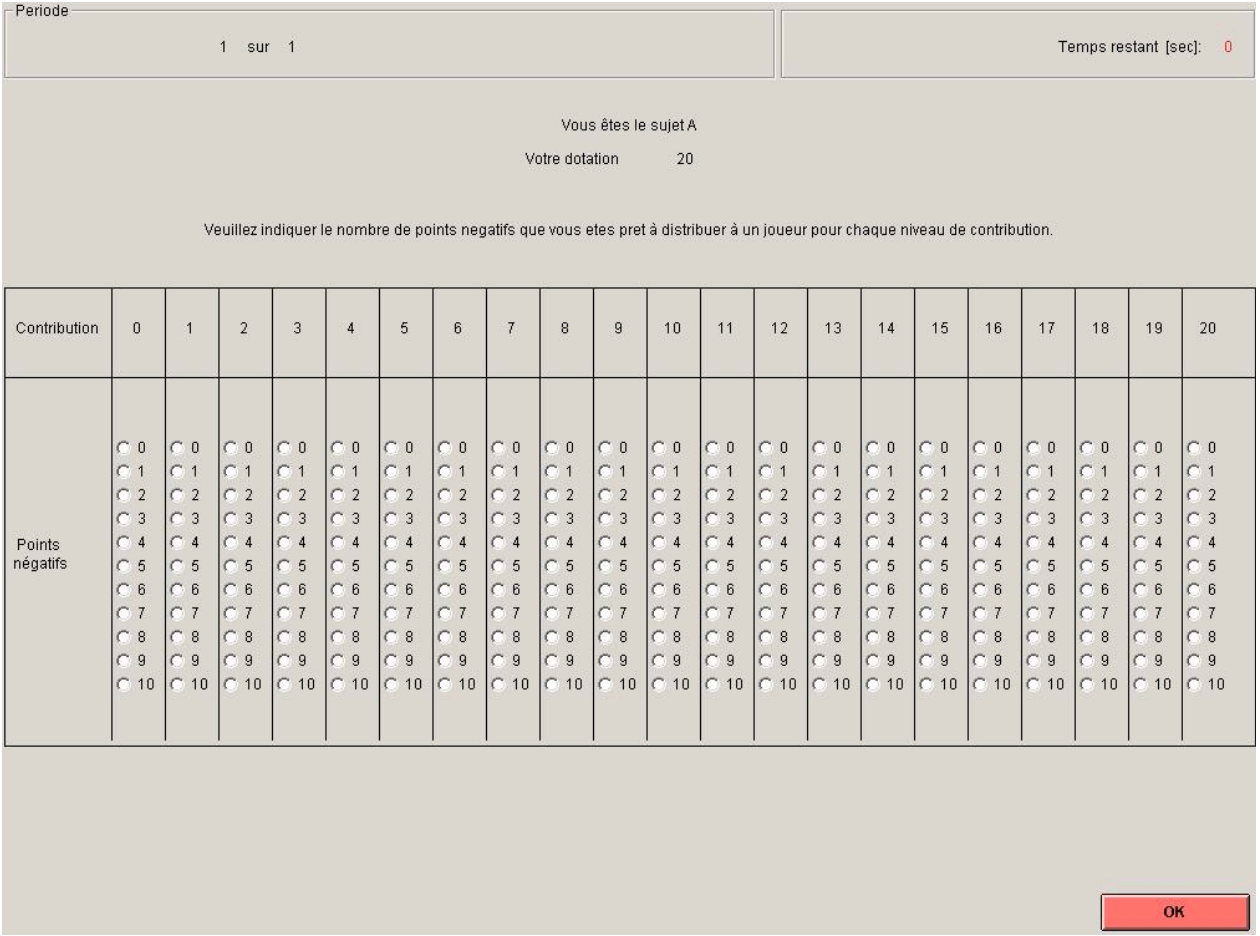
First stage

You announce the number of points you would like to assign to each other group member for each possible contribution level (between 0 and 20 ECU) to the project in the second stage. **The number of points you announce for a group member indicates your degree of disapproval for each contribution level (from 10 points for the highest disapproval to 0 point for no disapproval).** Your three other group members are informed of your announcement before they decide on their contribution level.

For the moment, the negative points you announce affect neither your payoffs nor the payoffs of your group members. They simply indicate to the others your willingness to reduce their payoffs for each possible contribution amount. It is only after every group member will have decided his contribution during the second stage that you will, in the third stage, confirm or modify your announced number of points. These points will then affect both your payoffs and the payoff of your group members, as indicated below.

- You announce the number of points that you would be willing to assign for each possible contribution level of your group members. You must enter a number, between 0 and 10, for each possible contribution. If you do not want express disapproval, you must enter 0.
- At the end of the first stage, the number of negative points you would be willing to assign for each contribution level will be announced to your group members. You are also informed on the total number of points that your three group members are willing to assign to you in the third stage for each of your possible contribution levels.

Below is the screenshot for the first stage.



Periode

1 sur 1

Temps restant [sec]: 0

Vous êtes le sujet A

Votre dotation 20

Veuillez indiquer le nombre de points negatifs que vous etes pret à distribuer à un joueur pour chaque niveau de contribution.

Contribution	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Points négatifs	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10	○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Second stage

You receive an endowment of 20 ECU. After being informed on the total number of points that you are susceptible to receive from the other group members for each possible contribution level, you decide on your contribution to the project.

You as well as the three group members decide simultaneously how much of your endowment you will allocate to the project, by indicating a number between 0 and 20. To validate your choice, click the OK button.

After all group members have made their decision, your screen will show you the total amount of ECU contributed to the project by the group members (including your contribution). You are also informed on your payoff for this stage.

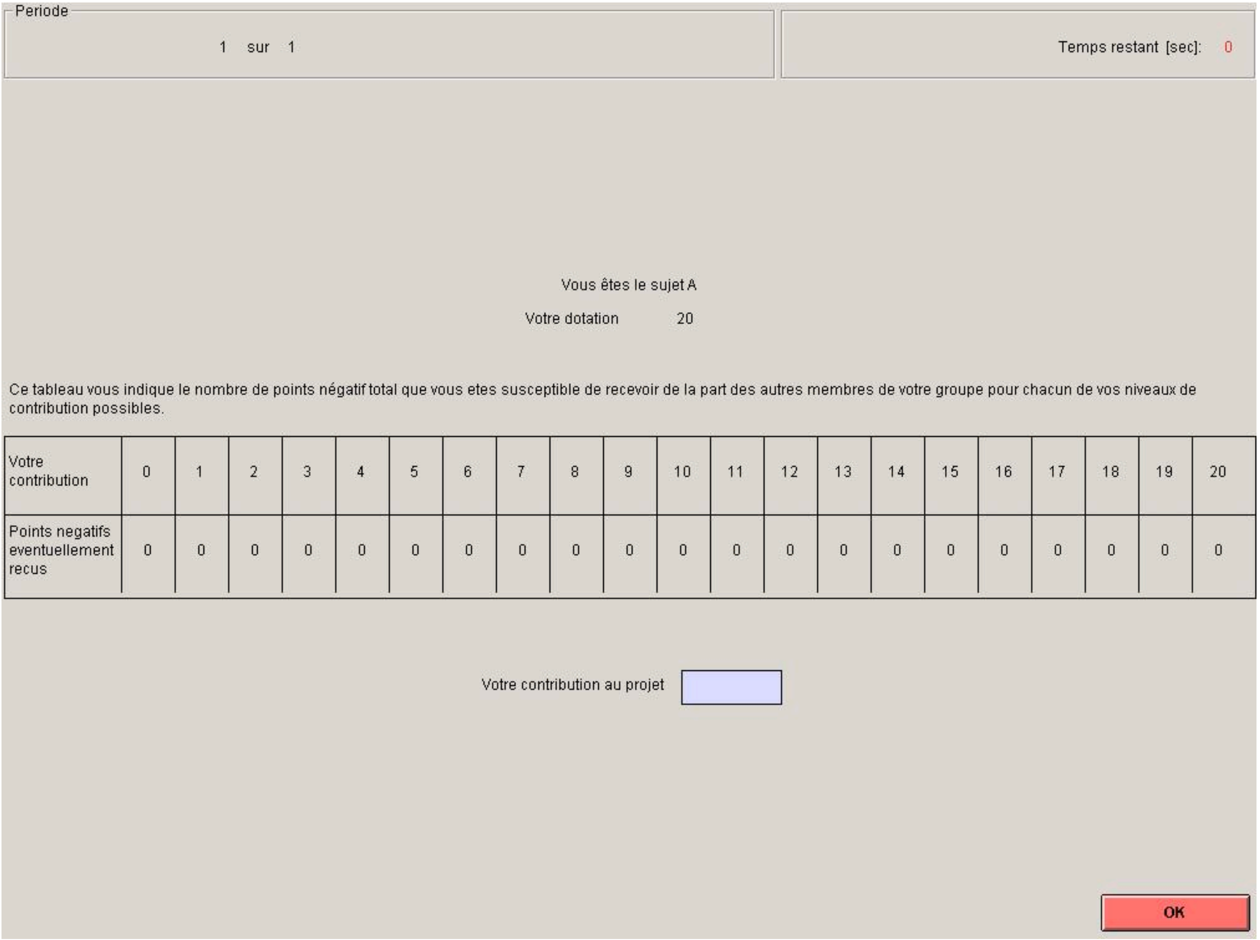
Your payoff in this second stage consists of two parts:

- the amount of your endowment which you have kept for yourself (i.e. 20 – your contribution to the project),
- the income from the project: this income represents 40% of the total contribution of all four group members to the project .

Your payoff in ECU in this second stage is computed by the program as follows:

(20-your contribution to the project) + 40%*(total contributions of the group to the project)

Below is the screenshot for the second stage.



Periode

1 sur 1

Temps restant [sec]: 0

Vous êtes le sujet A

Votre dotation 20

Ce tableau vous indique le nombre de points négatif total que vous etes susceptible de recevoir de la part des autres membres de votre groupe pour chacun de vos niveaux de contribution possibles.

Votre contribution	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Points negatifs eventuellement recus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Votre contribution au projet

The payoff of each group member is calculated in the same way, which means that each group member receives the same income from the project.

Suppose the total of the contributions of all group members is 60 ECU. In this example each member of the group receives a second-stage payoff from the project of 40% (of 60 ECU) = 24 ECU. If the total contribution to the project is 9 ECU, then each member of the group receives 40% (of 9 ECU) = 3.6 ECU from the project.

For each ECU of your endowment that you keep for yourself you earn an income of 1 ECU. Every ECU you contribute to the project instead increases the total contribution to the project by one ECU. The income from the project will increase by 0.4 ECU per person and so, the total income of the group from the project will rise by 1.6 ECU. This means that your contribution to the project also increases the income of the other group members.

On the other hand you will earn money from each ECU contributed by the other members to the project. For each ECU contributed by any group member you earn 40% (1) = 0.4 ECU.

Third stage

After being informed on the contribution of each of your group members, you can, if you like, reduce or leave unchanged their payoff by assigning points. **This number of points can be the same or different from the number you have announced in the first stage.** You can assign a particular number of points to a member of your group to express a level of disapproval (10 points for the highest disapproval, 0 points for no disapproval). Each point assigned to a particular group member reduces her second-stage income by two points.

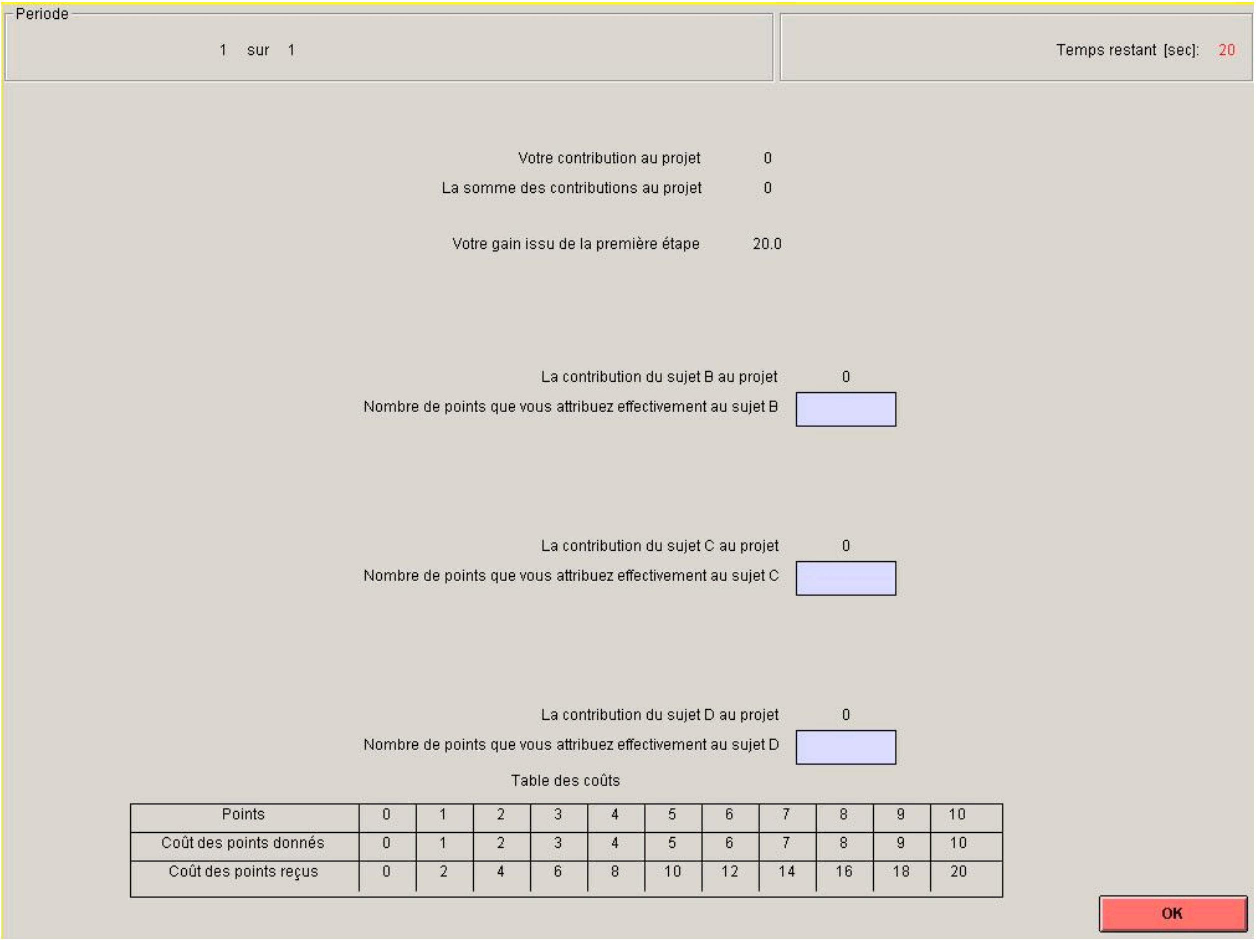
Your decision during the third stage depends on the actual contributions and can modify both your payoff and the payoff of your group members. Similarly, your payoff can be modified if the other group members wish to do so.

- You are informed of the contribution of each of your three group members to the project in the second stage of the game. Beware: the order in which each contribution is displayed is changed randomly in each period (in other words, for example the number that appears first on your screen does not always correspond to the decision of the same player).

- You decide next on how many points to give to each of the other three group members to reduce their payoff or leave it unchanged. Each point assigned to a group member reduces his second-stage payoff by 2 ECU. If you assign 0 point to another member, you do not modify his second-stage payoff. If you assign 1 point to a group member, you reduce his second-stage payoff by 2 ECU; if you assign 2 points, you reduce his second-stage payoff by 4 ECU; etc. You must enter a value for each member, between 0 and 10 points. If you do not wish to reduce the payoff of a specific member, then you must enter 0.

- If you assign points, you bear a cost that depends on the number of points you assign to each subject. Each point you assign reduces your second-stage payoff by 1 ECU. Your total cost is equal to the sum of the costs of assigning points to each of the other three group members. If you assign two points to one group member, this will cost you 2 ECU; if you assign 9 points to another member, this will cost you 9 ECU more; if you give the last group member no point, this will not cost you anything. In this example, the total cost of the assigned points is 11 ECU (2+9+0). These costs will be displayed on your screen. You can modify your decisions until you click the OK button.

Below is the screenshot for the third stage.



Periode

1 sur 1

Temps restant [sec]: 20

Votre contribution au projet 0

La somme des contributions au projet 0

Votre gain issu de la première étape 20.0

La contribution du sujet B au projet 0

Nombre de points que vous attribuez effectivement au sujet B

La contribution du sujet C au projet 0

Nombre de points que vous attribuez effectivement au sujet C

La contribution du sujet D au projet 0

Nombre de points que vous attribuez effectivement au sujet D

Table des coûts

Points	0	1	2	3	4	5	6	7	8	9	10
Coût des points donnés	0	1	2	3	4	5	6	7	8	9	10
Coût des points reçus	0	2	4	6	8	10	12	14	16	18	20

- Your final payoff in ECU in each period is calculated by the computer as follows:

Final payoff = (second stage payoff) - cost of received points in the third stage- cost of assigned points in the third stage

Note that in the calculation of payoffs, the cost of received points cannot exceed your second-stage income.

For example, if you received 3 points from the three other group members your second-stage payoff is reduced by 6 ECU. If you received 4 points, your second-stage payoff is reduced by 8 ECU. If you received 10 points, you lose 20 ECU of your second-stage payoff. You can possibly make a loss if you have assigned points. The amount of this loss corresponds to the cost of the points you have actually assigned to others.

Your third-stage payoff can therefore be negative if the cost of the points you have assigned exceeds your second-stage payoff net of the cost of received points. You can, however, avoid such losses with certainty through your own decisions.

To summarize

Each period consists of three stages.

- In the first stage, you announce the number of negative points you would be ready to assign to your group members for each possible contribution level. The three group members are informed on your announcement. Similarly, you are informed on the total numbers of points announced by your three other group members for each possible contribution.

- In the second stage, you choose your contribution to the project.

- In the third stage, you are informed on the individual contribution of each member of your group. You can assign negative points that will reduce their payoff and that can differ or not from your announcement in stage 1.

At the end of each period, the next period starts automatically. You receive a new endowment of 20 ECU.

Thank you for answering the questionnaire that has been distributed; we will check your answers individually. If you have any questions about these instructions, please raise your hand. We will answer your questions in private.

Communicating with the other participants during the experiment is strictly forbidden at the risk of being excluded from the session and from receiving your payment.