GROUPE D'ANALYSE ET DE THÉORIE ÉCONOMIQUE LYON - ST ÉTIENNE

GATE Groupe d'Analyse et de Théorie Économique Lyon Saint-Etienne

W P 1113

Emotions, Sanctions and Cooperation

Mateus Joffily, David Masclet, Charles N. Noussair, Marie-Claire Villeval

Mars 2011

Documents de travail | Working Papers

GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne

93, chemin des Mouilles 69130 Ecully – France
Tel. +33 (0)4 72 86 60 60
Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France
Tel. +33 (0)4 77 42 19 60
Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : gate@gate.cnrs.fr
Téléchargement / Download : http://www.gate.cnrs.fr – Publications / Working Papers

Emotions, Sanctions and Cooperation

Mateus Joffily [1], David Masclet [2], Charles N. Noussair [3], Marie Claire Villeval [4*]

February 2011

Abstract: We use skin conductance responses and self-reports of hedonic valence to study the emotional basis of cooperation and punishment in a social dilemma. Emotional reaction to free-riding incites individuals to apply sanctions when they are available. The application of sanctions activates a "virtuous emotional circle" that accompanies cooperation. Emotionally aroused cooperators relieve negative emotions when they punish free riders. In response, the free-riders experience negative emotions when punished, and increase their subsequent level of cooperation. The outcome is an increased level of contribution that becomes the new standard or norm. For a given contribution level, individuals attain higher levels of satisfaction when sanctioning institutions are in place.

Keywords: Emotions, Sanctions, Cooperation, Experiment, Skin Conductance Responses

JEL-Codes: C92, D62, D63, D64, D74

[1] IPUB, Federal University of Rio de Janeiro, Av. Venceslau Brás, 71, 22290-140, Rio de Janeiro, Brazil; CNRS, Centre des Neurosciences Cognitives, Bron, France. E-mail: joffily@isc.cnrs.fr.

[2] CNRS, CREM, 7 Place Hoche, 35065 Rennes, France, CIRANO, Montreal. Email: david.masclet@univ-rennes1.fr.

[3] Department of Economics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: C.N.Noussair@uvt.nl.

[4] University of Lyon, Lyon, F-69007; CNRS, GATE, 93, Chemin des Mouilles, F-69130, Ecully, France. E-mail: villeval@gate.cnrs.fr.

* All authors contributed equally to this work. They thank S. Ferriol for programming assistance and K. Straznicka and J. Rosaz for valuable research assistance. We thank Enst Fehr, Astrid Hopfensitz, Andrew Oswald, Frans van Winden, Lise Vesterlund, and participants at seminars at the University of Warwick, at the University of Pittsburgh, at Tilburg University, and at the ASFEE workshop on "How can Neurosciences inform Economics?", Grenoble. A grant from the Rhone-Alpes Region (CIBLE program) is gratefully acknowledged.

1. INTRODUCTION

Understanding how cooperation in groups arises is a longstanding focus of research in the social sciences. A number of scholars have argued that the existence of punishment opportunities aids in creating and sustaining cooperation in social dilemmas (Homans, 1961; Blau, 1964; Coleman, 1990; Elster, 1998; Bowles and Gintis, 2001). Behavioral experiments have supported this proposition (Yamagishi, 2006; Ostrom *et al*., 1992; Fehr and Gaechter, 2000, 2002). While costly sanctions have a detrimental direct effect on overall welfare because they waste resources (Houser *et al.*, 2008; Milinski *et al.,* 2008; Herrmann *et al*., 2008), in the long run the availability of costly punishment increases surplus through its strong positive effect on cooperation (Gaechter *et al*., 2008). This is especially the case when punishment can operate, in conjunction with reputation, in a setting in which the same players interact repeatedly (Milinski *et al*., 2006; Rand *et al*., 2009).

Game theorists have shown that there are many settings in which cooperation, as well as punishment that is costly to sanctioners, can occur as equilibrium behavior on the part of selfish agents (Fudenberg and Tirole, 1991). However, the experiments listed above show that cooperation and the application of punishment occur under more general conditions than those consistent with traditional game theory. This raises the possibility that the propensities to cooperate and to apply costly punishment may also have an emotional foundation. It is known that emotional processes are involved in the decision to punish in two-person interactions. In particular, a sentiment of anger accompanies the application of costly punishment (Bosman and van Winden, 2002; Ben Shakhar *et al*., 2007; Hopfensitz and Reuben, 2009). It has also been shown that when observing opportunistic behavior, anterior insula activation, which is typically associated with aversive stimuli, correlates with subsequent individuals' decision to punish others (Sanfey *et al*., 2003). Punishment of social norm violators has also been related to satisfaction, as punishment activates the dorsal striatum, a brain area often associated with pleasant stimuli and reward-driven actions (de Quervain *et al*., 2004). Emotional processes are also involved in the decision to mutually cooperate. For example, striatum activation, typically correlated with reward, is also associated with mutually cooperative behavior in prisoner's dilemma games (Rilling *et al.,* 2002; Rilling *et al*., 2004).

In the research reported here, we investigate the connection between emotions, punishment and cooperation. We measure the emotional states of participants in a social dilemma at various key moments of their interaction. Specifically, we study the emotional responses related to the level of cooperation an individual exhibits, as well as to the observation of others' cooperation levels. When sanctions are possible, we consider the emotions that accompany the decision to punish and the severity of the punishment applied, as well as the receipt of punishment. We also consider the relationship between the emotional response to sanctions received and subsequent changes in cooperation. Measuring emotions at these various moments allows us to identify the link between emotions and cooperation and its role in the emergence of social norms in groups.

Our focus is on the dynamics of the emotional profile of group members, its relationship to the emergence of either a free-riding norm or a cooperative norm within a group, and how it interacts with the existence of a punishment institution. We measure emotional state along two dimensions: (i) the level of arousal, and (ii) the valence, pleasant or unpleasant. The Skin Conductance Response (SCR) of individuals is used as a metric of the emotional arousal. Self-reports of emotional state, on a Likert scale, are used to measure the valence of the emotional experience. The SCR is a relatively simple measurement to perform in a multi-subject experimental setting, and allows for a straightforward interpretation of the data in terms of the intensity of emotional arousal. Unfortunately, SCR is ambiguous with regard to the valence of emotion, that is, whether the emotional state is positive or negative. Self-reports of experienced emotion are very informative about whether an individual views her emotional state as positive or negative. However, self-reported magnitudes may be unreliable indicators of intensity.

2. THE EXPERIMENT

2.1. The experimental setting, parameters, and procedures

We examine the role of emotions in the emergence and maintenance of cooperation with a paradigm referred to as the Voluntary Contributions Mechanism (see for example Marwell and Ames, 1979). This setting is widely used in the experimental investigation of cooperation in social sciences. A group of individuals are each presented with an opportunity to allocate an endowment between two uses. The first use is a private account, which benefits only the individual. The second is a group account, which benefits all group members. The payoffs are specified so that it is a dominant strategy for each individual to place his entire endowment in his private account, but attaining the social optimum requires all individuals to allocate their full endowment to the group account. The percentage of endowment allocated to the group account provides a measure of the level of cooperation that the group exhibits.

In our experiment there were two treatments, called Baseline and Sanction (see instructions in Appendix A). The design was based on that of developed in Fehr and Gaechter (2000). Only one treatment was in effect in a given session, and no subject participated in more than one session. The twelve participants that attended each session were assigned to groups of four with fixed membership. The four members of each group interacted repeatedly but anonymously for 20 periods under a "partner" matching protocol.

A session of the Baseline treatment proceeded in the following manner. In each period, each participant received an initial endowment of 20 ECU (*Experimental Currency Units,* 100 ECU = 2 Euro). Players then simultaneously chose an amount between 0 and 20 ECU to contribute to their group account, and the remainder would be placed in their private account. Within each group, contributions were totaled, multiplied by 1.6, and redistributed equally among the members of the group. This meant that each group member received 40% of the total amount the group assigned to the group account, in addition to the amount he assigned to his own private account. Since each ECU kept in his private account yielded an individual 1 ECU, whereas his return from each ECU contributed to the project was 0.4 ECU, a player's dominant strategy was to allocate his entire endowment to his group account. However, the maximum feasible group payoff would be attained only if each player contributed his full endowment to the group account. This would yield each group member 32 ECU. The average level of contribution to the group account was taken as a measure of cooperation.

In the Sanction treatment, a second stage was added to each period, in which participants could punish the other members of their group. After being informed of the contribution of each group member, a participant could assign between 0 and 10 punishment points to each member. Sanctioning was costly to both the punisher and his target. Each point assigned to an individual reduced the recipient's earnings by 10 percent, with a maximum possible reduction of 100 percent. The cost of punishment points for the participant who assigned them was convex in the number of points. The cost function for the assignment of punishment is given in Table 1.

Table 1. The cost of sanctions to the punisher and to the target

Number of punishment points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher	0	1	2	4	6	9	12	16	20	25	30
Percentage reduction of target's payoff	0	10	20	30	40	50	60	70	80	90	100

Sanctioning decisions were made simultaneously and anonymously, so that it was not possible for a participant to identify who had punished her. Classical game-theoretic reasoning produces a unique subgame-perfect equilibrium to the two-stage game of the Sanction treatment. Since punishment is costly, no group member ever uses the opportunity to punish, for any history of play. Thus, facing a threat that is not credible, each player places her entire endowment in her own private account.

The experiment consisted of four sessions. These sessions were conducted at the *Groupe d'Analyse et de Theorie Economique* (GATE), Lyon, France. 48 subjects (of whom 41.67% were males) were recruited by means of the ORSEE software (Greiner, 2004) from undergraduate courses in the local business and engineering schools. The experiment was computerized using the REGATE program developed at GATE (Zeiliger, 2000). Because there were four individuals in each group, and group membership remained the same for the entire experiment, there were twelve independent observations at the group level. There were six observations in each treatment. At the end of the session, the participants were required to complete a demographic questionnaire and then they were allowed to leave the laboratory. They were informed at the beginning of their session that a person who is not aware of the content of the experiment would disburse their earnings. [1]

2.2. Skin Conductance Response acquisition and analysis, and valence elicitation

Multivariate analyses of verbal reports of emotional stimuli (see for example Mehrabian and Russell, 1974) have shown that most of the variance in emotional reports can be explained by two main factors: valence (varying from negative to positive) and arousal (varying from low to high). This two-dimensional structure of reported emotions is mediated by appetitive and defensive motivational brain systems (Lang et al., 1992). More recently, research has begun to identify the brain structures underlying these two motivational systems (Anders et al., 2004). Self-reports of the emotional valence and the arousal dimensions are correlated with autonomic and somatic responses to emotional stimuli (Bradley et al., 2001). More specifically, facial electromyography, heart rate and startle reflex are correlated with the emotional valence dimension, whereas skin conductance responses positively correlate with emotional arousal independently of valence.

In this study, we interpret skin conductance responses as a measure of physiological arousal associated with an emotional state. During our sessions, skin conductance was

[1] Upon arrival, each subject drew a tag indicating their designated computer. Next, after washing his hands with a neutral soap, each subject was allowed to enter the laboratory and sit in front of his computer. Then, an assistant put electrodes on the non-dominant hand. Participants were required to keep the electrodes on until the end of the session. After checking the quality of the signal recording and verifying that all participants were connected, the instructions for the experiment were distributed and read aloud. Understanding of the rules of the game was checked with a questionnaire. Subjects' questions were answered in private.

simultaneously recorded for each of the 12 participants present. The sessions took place in a noiseless laboratory with stable temperature set to 21^{o} C. Skin conductance was recorded with a BIOPAC MP150W system and two TEL100C telemetry modules (BIOPAC Systems, EU). Two Ag/AgCl electrodes filled with 0.5% saline in a neutral base paste were placed on the subject's distal phalanges of the middle and the index fingers of the non-dominant hand, after the attachment site had been cleaned with a neutral soap (Dawson et al., 2000). A constant voltage of 0.5V was applied between the electrodes. The skin conductance signal was amplified (x2000) and low-pass filtered (30Hz) before being sampled at 125Hz. Skin conductance activity was continuously recorded until the end of the session.

Skin conductance was analyzed in response to four events of interest. (a) The first is at the time of the contribution decision. Each participant was allowed to make her contribution decisions at her own pace. After the last participant had made his decision, an 8 s interval was imposed before all participants were requested to report the valence of their feelings regarding their own contribution decision. (b) The second event of interest is the time of receipt of information about each other group member's contribution. This information was displayed on the screen for 8 s, and participants were simultaneously requested to report the valence of their feelings toward each other player's contribution. There were two more events of interest in the Sanction treatment, during which SCR was analyzed. The first of these was (c), when participants decided whether and how much to punish other members of their group. The punishment decision was also self-paced. An interval of 8 s was imposed between the decision time of the last participant and the request to all participants to report their feelings regarding their own punishment decisions. Finally, (d) the punishment received from others was displayed on each participant's screen for 8 s, and it was followed by a request to participants to report the valence of their feelings toward the total sanction they received.

The participants were required to report how they felt on a 10-point Likert scale of hedonic valence, which ranged from "extremely unpleasant" to "extremely pleasant". A level of 5 is interpreted as a neutral sentiment, since 0 is the most unpleasant, and 10 is the most pleasant possible evaluation. As indicated above, this reporting occurred just after each event (a) – (d). When indicating a reaction to others' contribution levels at time (b), subjects were asked to submit a separate reaction to each of the three other players' contributions. Similarly, at time (c), a separate reaction was reported for each assignment of points to the three other group members.

Skin conductance responses were analyzed within specific time windows corresponding to events (a)-(d). The analysis windows started 1 s after the onset of each event. The purpose of the lag was to account for the SCR's latency. The analysis window's duration varied by event.

Events (a) and (c) had variable analysis window durations. These are labeled as T1 and T2, respectively, in Figure 1. Events (b) and (d) had fixed durations of 7 s and 2 s, respectively, to comply with the amount of information displayed on the screen at each moment.

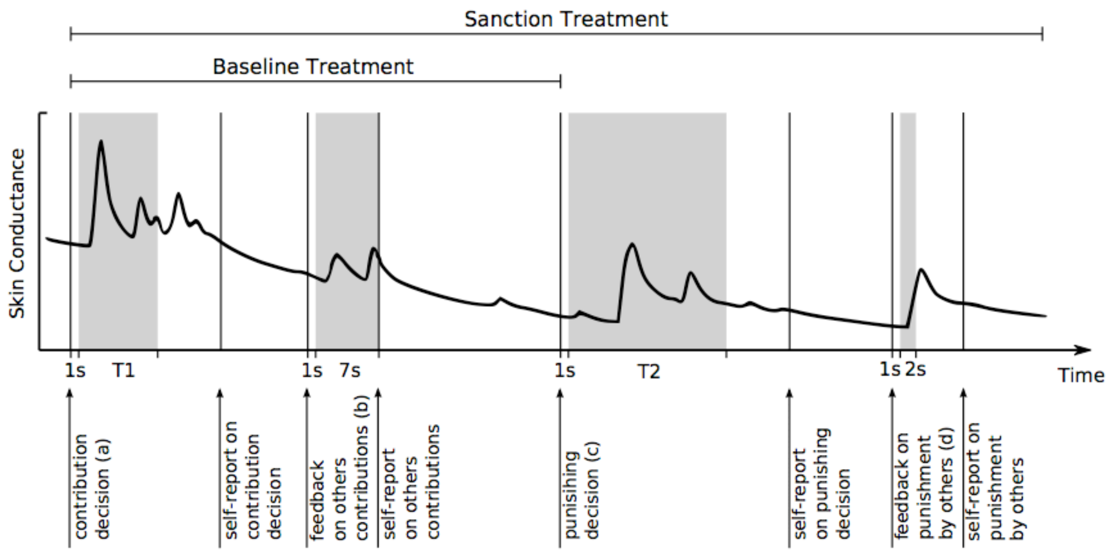

Fig. 1. Timing of Events in the Baseline and Sanction Treatments within a Period[2]

The raw data were preprocessed in the following manner. The skin conductance signal was low-pass filtered at 0.5Hz offline, using a 5th order Butterworth low-pass digital filter. The onset and peak of the skin conductance responses (SCR) were automatically detected when the first derivative of the filtered signal changed sign with a routine written in Matlab (The MathWorks Inc., USA). Onsets were identified by a negative to positive zero crossing, while peaks were identified by a positive to negative zero crossing. The SCR amplitude was calculated as the difference between the signal amplitude at the peak and the onset times. The SCR amplitude was thresholded at 0.02 μS (*5*). The whole signal was visually inspected prior to further analysis and ectopic response was removed. The detection of an accelerative deflection during the interval between onset and peak times indicated overlapping SCRs during rise time. In that case, the two SCR were *i*) separated, if they could be related to different moments of interest,

[2] A typical time path of skin conductance signal from a single subject within a period, and the segments extracted for analysis, are illustrated. Four events of interest were analyzed: (a) participant's own contribution decision, (b) receipt of feedback about each other group member's contribution, (c) participant's own punishment decision and (d) receipt of feedback about the sanctions received from others. The windows corresponding to events (a) and (c) had a variable duration, and those corresponding to (b) and (d) had a fixed duration of 8s. Shaded grey areas indicate onset and duration of the SCR analysis windows. The analysis windows at moments (a) and (c) have variable durations T1 and T2, respectively, due to self-paced decision-making. A minimum interval of 8 s was imposed between T1 (T2) and self-reports of the contribution (punishment) decision.

or *ii*) summed together, if they were related to the same moment (Boucsein, 1992). In the case of overlapping responses during recovery-time, the amplitude scoring based on the difference between the signal amplitude at the peak and the onset times is sufficiently accurate and it is used as a standard procedure (Edelberg, 1967). In order to minimize SCR overlap between events of interest, we imposed a minimum interval between events of 8 seconds.

3. RESULTS

3.1 Contribution behavior and treatment differences

We first consider whether measuring SCR and eliciting self-reports affected behavior. We do so by verifying that in our experiment, the same qualitative patterns are observed as in previous studies. Figures 2 and 3 display the average individual contribution over time in the Baseline and Sanction treatments respectively, for each group. A value of 20 is the maximum possible, and corresponds to the social optimum with full cooperation. A value of 0 corresponds to the Nash equilibrium with zero cooperation.

Figure 2 shows a high dispersion of contributions across groups in the Baseline treatment, but typically with positive levels of cooperation in early periods, and with a decline over time to levels close to zero by the end of the sessions. Figure 3 shows that in the Sanction treatment, cooperation increases over time to close to 100% by the end of the sessions. These patterns are consistent with previous studies (Fehr and Gaechter, 2000; Masclet *et al.*, 2003).

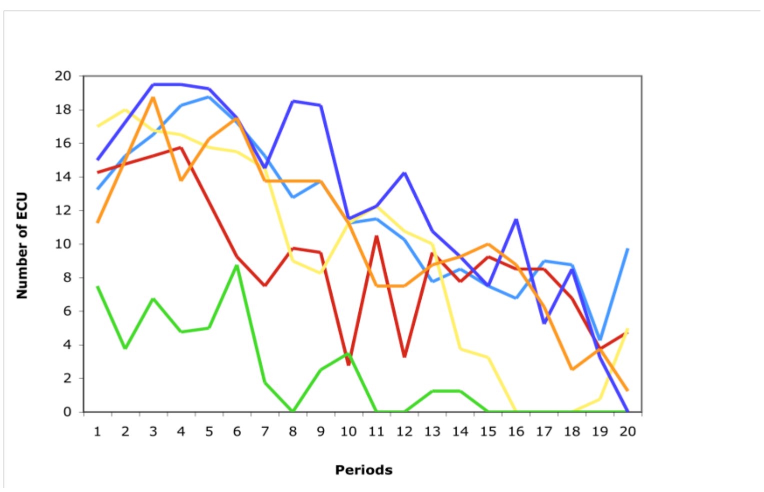

Fig.2. Average contribution by group in the Baseline treatment

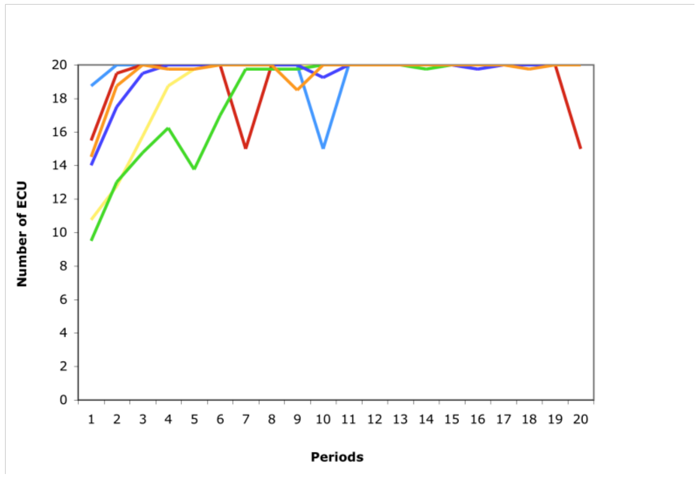

Fig.3. Average contribution by group in the Sanction treatment

In the Baseline treatment, the average contribution is 13.04 ECU (S.D. = 3.09) in period 1, 8.58 (S.D. = 3.95) in period 10, and 3.46 (S.D. = 3.55) in period 20. In contrast, in the Sanction treatment, average contributions converge to near the social optimum of full contribution of 20 units after a few periods. The average contribution is 13.83 ECU (S.D. = 0.43) in period 1, 19.04 (S.D. = 0.22) in period 10, and 19.17 (S.D. = 0.21) in period 20. The difference between the two treatments is not significant in period 1 (Mann-Whitney tests, M-W hereafter, two-sided, p = 0.749 when using each group as a unit of observation, and p = 0.790 when using each individual as an observation). However, the difference is highly significant in period 10 (M-W, p = 0.003, each group as an observation) and remains so in period 20 (M-W, p = 0.003). In the rest of the paper all of the M-W tests are conducted at the group level and are two-sided, unless specified otherwise.

3.2 Emotions at the time of the contribution decision

SCR magnitudes are greater for individuals who cooperate less, in both treatments. The data are shown in the upper panel of figure 4. Contributions are grouped into four categories, based on how many of the 20 tokens were allocated to the group account. The top panel relates the subject's contribution level and his average SCR magnitude when contributing. The SCR magnitude is the mean SCR value computed across all trials, including those in which no significant response occurred. The figure shows a negative correlation between contribution and arousal in the Baseline treatment. In contrast, the arousal level is similar for different contribution levels in the Sanction treatment, though it is much greater for the small number of observations (six), in which five or fewer tokens were contributed.

We also estimated a random-effects Tobit model in which the dependent variable is the SCR magnitude the subject exhibits when making his contribution decision. The estimation is shown as model 1 in Table 2. The Tobit specification is used because of the left-censored dependent variable. Subject-specific random effects account for the fact that the same subjects make repeated decisions. The independent variables include the subject's contribution, a dummy variable for the Sanction treatment, a time trend captured by the period number, and the subject's gender. The estimates (log-likelihood = -613.741, Wald-Chi^2 = 30.27, p<0.001, N = 960) indicate that there is a significant negative correlation between the contribution an individual makes and the SCR magnitude he exhibits (coeff. = -0.007, p = 0.014). The two treatments generate a similar level of arousal at the time of the contribution decision as there is no significant difference by treatment (p = 0.506). Individuals feel less aroused over time, which is likely due to a habituation effect (coeff. = -0.014, p<0.001).

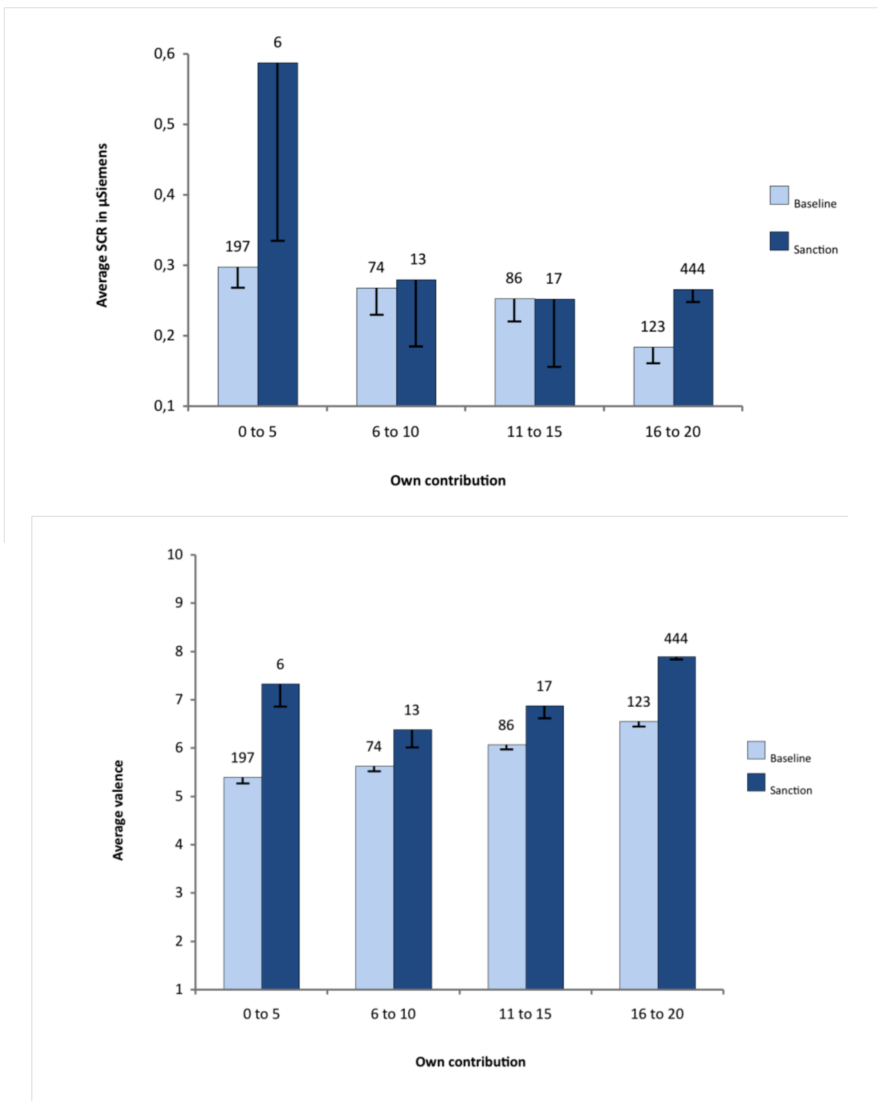

Fig.4. Physiological arousal (SCR, upper panel) and valence of emotions (lower panel) at the time of submission of contribution decisions

Table 2. Determinants of the SCR and valence of emotions when subjects contribute[3]

Dependent variable:	SCR when contributing Random-effects Tobit (1)	Self-reported valence Ordered Probit (2)
Contribution	-0.007 ** (0.003)	0.029** (0.014)
Sanction treatment	0.060 (0.090)	0.622*** (0.182)
Time trend (t)	- 0.014*** (0.003)	0.003 (0.016)
Male	0.017 (0.087)	0.105 (0.219)
Constant	0.359*** (0.087)	-
N	960	960
Left-censored obs.	313	-
Log-likelihood (pseudo)	-613.741	-1940.607
Wald χ^2	30.27	-
$p>\chi^2$	0.000	-
Pseudo R^2	-	0.046

Note: ** and *** indicate statistical significance at the 0.05 and 0.01 level, respectively. Robust standard errors are in parentheses.

The bottom panel of Figure 4 displays, for each contribution level, the average self-reported emotional state. A value of 0 corresponds to “extremely unpleasant” and 10 to “extremely pleasant.” In the Baseline treatment, the average self-reported valence is 5.40 (S.D. = 3.09) for contributions less than 6 ECU, 5.87 (S.D. = 1.58) for contributions between 6 and 15 ECU and 6.56 (S.D. = 2.18) for contributions above 15 ECU. In the Sanction treatment, the corresponding mean values are 7.33 (S.D. 2.16), 6.67 (2.14), and 7.90 (S.D. = 2.11). In the Sanction treatment, if one excludes as outliers the 6 observations with contributions below 6 ECU, then the more group members cooperate, the more positive the feelings they report. Interestingly, individuals feel more positive on average in an environment where sanctions are available than in a setting with no possibility of sanctions (Baseline treatment: mean = 5.85, S.D. = 2.48; Sanction treatment: mean = 7.81, S.D. = 2.13). This is the case, even controlling for the greater contributions under Sanction, as is suggested in figure 4.

[3] Model (1) in Table 2 displays the results of the estimation of a random-effects Tobit model. In all of the Tobit regressions in this paper, we employ a similar random effects structure. We also estimate a Tobit model with clustering of standard errors at the group level, but without random effects. These are not reported here, but the estimates are qualitatively similar for all of the results reported here. Model (2) displays the results of the estimation of an ordered Probit model, with robust standard errors and clustering at the group level, in which the dependent variable is the self-reported valence of emotions when contributing. All Probit regressions in this paper use the same error and clustering assumptions. In all regressions, ** and *** indicate statistical significance at the 0.05, and 0.01 level, respectively. Standard errors are in parentheses.

The estimates of ordered Probit model, in which the dependent variable is the self-reported emotions at the time of contribution, are shown as model (2) in Table 2. The estimates show that the more one contributes, the more one's feelings are positive (coeff. = 0.029, $p = 0.043$). They confirm that emotions are more positive in the Sanction treatment than in the Baseline (coeff. = 0.622, $p = 0.001$), controlling for the other variables in the specification. In contrast with the SCR measures, there is no significant time trend in the direction of emotions ($p = 0.867$).

To summarize, these findings suggest that participants in this social dilemma face a trade-off between monetary incentives to free ride and valence of the emotional state. The more a subject contributes, the lower is her monetary payoff in the current period but the more positive is her self-reported emotional state. Furthermore, greater arousal is associated with lower contributions, suggesting the presence of a substantial emotional cost to free-riding – since valence is more negative - for the average individual.

3.3 The emotional impact of observing other group members' contributions

Figures 5 and 6 show the relationship between the contributions of other group members and one's own emotional response. Figure 5 relates the average SCR of an individual to the average contribution in the group (excluding the individual's own contribution). In the figure, each bar indicates the average response in reaction to each level of difference between the average and *i*'s contributions. For example, the (-20 to -6) category on the left corresponds to the cases when the average contribution is at least 6 ECU less than the subject's own contribution. In contrast with the SCR measures, we can relate self-reports to each other group member's contribution.

Figure 5 shows that in the Baseline treatment, Skin Conductance Response is greater when the individual learns that he has contributed less than the average (as shown in the positive range on the right portion of the figure) than when he learns that he has contributed more. In the Sanction treatment, learning that one has contributed more than the average triggers greater arousal, presumably because cooperators are upset of discovering free-riding by others in an environment where the social norm is to contribute. For the few observations in which the individual's own contribution is much lower than the average, there is also a strong arousal response, presumably due to the anticipation of receiving sanctions. In general, skin conductance response is greater in the Sanction than in the Baseline treatment.

A random-effects Tobit model is also estimated, in which the SCR magnitude, registered at the time of receipt of feedback about others' contributions, is the dependent variable. The results are shown in column (1) of Table 3. The independent variables include the mean contribution of other group members, the absolute values of positive and of negative differences between one's own contribution and the average contribution of other group members (a negative difference means that one contributes less than the average), a dummy variable for the Sanction treatment, a time trend, and a dummy variable for the individual's gender.

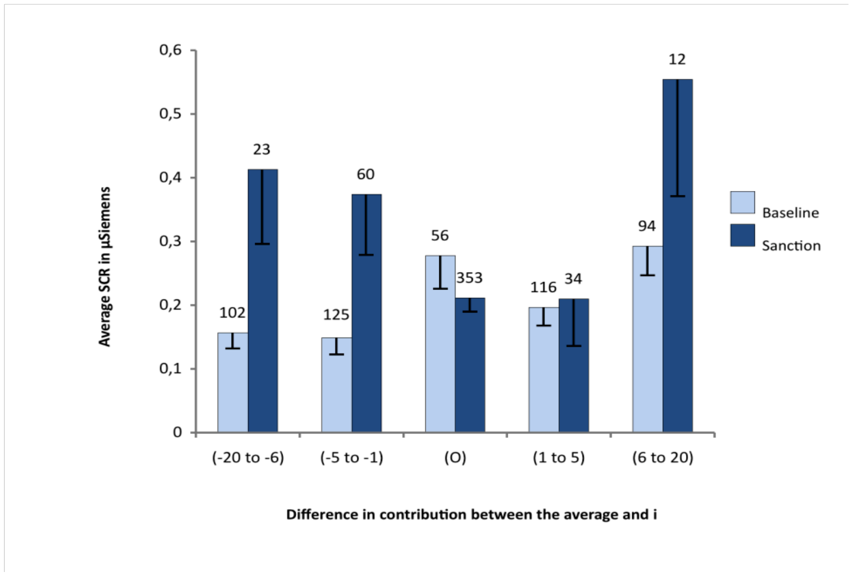

Fig.5. Average SCR magnitude when being informed of others' contributions as a function of the difference between the average contribution of others and one's own contribution, $(\bar{c}_{-i} - c_i)$

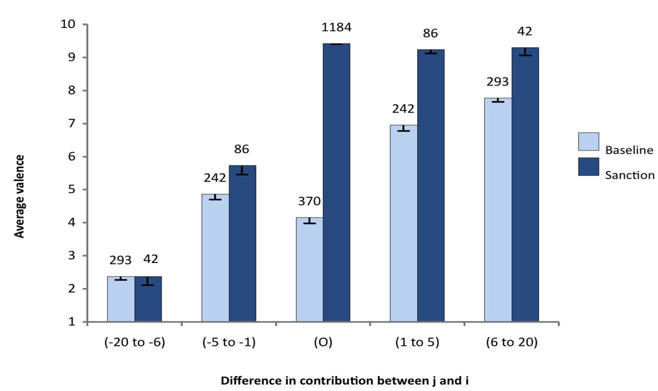

Fig.6. Average self-reported valence of emotions when being informed of another player's contribution, as a function of the difference between the other's contribution and one's own contribution $(c_j - c_i)$

Table 3. Determinants of SCR and valence of emotions of player I when receiving feedback about others' contributions

Dependent variables:	SCR when receiving feedback - Random-effects Tobit (1)	Valence when receiving feedback Ordered Probit (2)
Mean contribution of other group members $\overline{c}_{-i}$	-0.016 ***(0.006)	-
Contribution of the other group member c_j	-	0.042*** (0.006)
Absolute value of positive difference from the average: $(\max\{0, c_i - \overline{c}_{-i}\})$	0.008 (0.008)	-
Absolute value of positive difference from own contribution $(\max\{0, c_j - c_i\})$	-	0.062*** (0.013)
Absolute value of negative difference from the average: $(\max\{0, \overline{c}_{-i} - c_i\})$	0.025*** (0.006)	-
Absolute value of negative difference from own contribution $(\max\{0, c_i - c_j\})$	-	-0.123*** (0.026)
Sanction treatment	0.286*** (0.108)	0.558*** (0.177)
Time trend (t)	-0.016*** (0.004)	-0.015 (0.237)
Male	0.118 (0.093)	-0.149 (0.170)
Constant	0.146 (0.123)	-
N	960	2880
Left-censored obs.	414	-
Log-likelihood / Log pseudo-likelihood	-748.290	-4132.843
Wald χ^2	37.74	-
$p>\chi^2$	0.000	-
Pseudo R^2		0.246

Note: *** indicates statistical significance at the 0.01 level. Robust standard errors are in parentheses in model (2).

The estimates of the model (1) show that the physiological arousal decreases in the amount contributed by others (coeff. = -0.016, p = 0.007). In addition, the more a subject deviates negatively from the average behavior within the group, the more he is physiologically aroused (coeff. = 0.025, p<0.001). This shows that, controlling for the degree of cooperation in the group, the participant is physiologically aroused by learning about his comparative free-riding behavior. In contrast, positive own difference from the average of the group is not significant (p = 0.272). Physiological arousal is significantly higher in the Sanction treatment (coeff. = 0.286, p = 0.008). The time trend is negative and significant (coeff. = -0.016, p<0.001).

Separate random-effects Tobit regressions by treatment (not reported here but available upon request) show that, controlling for the mean contribution in the group, the coefficient of the absolute value of negative differences between one's own contribution and the average contribution of other group members is twice as high in the Sanction treatment (coeff. = 0.032, p = 0.040) than in the Baseline (coeff. = 0.016, p = 0.013). In addition, the time trend is not significant in the Baseline treatment (p = 0.442) but it is significantly negative in the Sanction treatment. This might be a consequence of the convergence of all groups toward full cooperation over time in this treatment.

Figure 6 shows that in both treatments, when informed of others' contributions, individuals experience more positive emotions if they learn that another player has contributed more than they. In general, there is greater satisfaction in the Sanction than in the Baseline treatment. We also conducted a regression analysis, reported in column (2) of Table 3, in which the self-reported hedonic valence when receiving feedback on others' contributions was the dependent variable. The regressions use an ordered Probit specification with robust standard errors and clustering at the group level. The independent variables include the other group member's contribution, the absolute value of the positive difference between the other group member's contribution and one's own contribution, the absolute value of the negative difference between the other group member's contribution and one's own contribution, a dummy variable for the Sanction treatment, a time trend, and the participant's gender.

The estimates show that the more another group member contributes, the more positive the self-reported valence in response (coeff. = 0.042, $p < 0.001$). A negative difference between the other group member's contribution and one's own contribution negatively affects the valence of emotions in comparison with a situation in which contributions are equal (coeff. = -0.123, $p<0.001$). In contrast, positive differences improve the valence of emotions (coeff. = 0.062, $p<0.001$). All else equal, the valence of emotions is more positive in the Sanction treatment (coeff. = 0.558, $p = 0.002$). There is no time trend ($p = 0.237$). Separate ordered Probit regressions by treatment (not reported here but available upon request) show that, controlling for the other's contribution, the coefficient of the absolute value of negative differences between one's own contribution and the other group member's contribution is twice as high in the Sanction treatment (coeff. = -0.124, $p<0.001$) than in the Baseline (coeff. = 0.056, $p<0.001$). In the Baseline, the valence of emotions becomes more and more negative over time as free riding increases (coeff. = -0.035, $p<0.001$), while no time trend is detected in the Sanction treatment ($p = 0.612$).

In summary, learning that others have been relatively cooperative triggers a positive emotional response. Thus, higher contributions yield both a positive financial and emotional externality on other players. The level of arousal is greater, the more one has contributed relative to others in the Baseline treatment. In the Sanction treatment, a different pattern exists: making contributions both higher and lower than others triggers high arousal. This may be related to the anticipation of assigning or receiving sanctions in the next stage of the game.

3.4 Emotions and the decision to punish

As shown by figure A in Appendix, there is a strong tendency for i to punish j more, the less that j contributed compared to i. There is also a modest tendency for i to punish j more, the more j has contributed compared to i. A small difference (where a negative number denotes how much more the sanctioner contributes than the target in the current period), between (-2,+2), triggers on average 0.04 punishment points. The corresponding number of points is 1.04 for the range (-3, -8), 1.56 for the range (-9,-14), and 2.43 for (-15, -20). This is a typical pattern in VCM experiments. This pattern, in conjunction with the emotional response experienced when being informed about the contributions of others, suggests that greater arousal and more negative emotional valence may lead to the assignment of punishment.

Table 4. Emotions and the decision to punish

Dependent variables:	Probability of sanctioning another group member RE Probit (1)	Severity of sanctions on a group member j RE Tobit (2)
Physiological arousal when contributing	-0.103 (0.243))[-0.017]	0.502** (0.241)
Physiological arousal when being informed of others' contributions	0.281* (0.159)[0.047]	0.113 (0.179)
Valence of emotions when contributing	0.076 (0.059) [0.013]	0.116* (0.063)
Valence of emotions when receiving feedback about others' contributions	-0.424*** (0.090) [-0.071]	-0.584*** (0.064)
Average contribution of others ($\overline{c}_{-i}$)	-0.088* (0.050) [-0.015]	0.054 (0.054)
Difference between i's and the average contribution of others (c_i - $\overline{c}_{-i}$)	0.047 (0.030) [0.008]	-0.024 (0.031)
Positive difference between j's and the average contribution (max{0, c_j - $\overline{c}$ })	-	0.061 (0.070)
Negative difference between j's and the average contribution (max{0, $\overline{c}$ - c_j })	-	0.088*** (0.033)
Time trend (t)	-0.069*** (0.019)[-0.012]	-0.100*** (0.024)
Male	-0.133 (0.269) [-0.022]	-0.184 (0.430)
Constant	4.311*** (0.806)	0.998 (1.001)
N	480	1440
Left-censored observations	-	1327
Log-likelihood	-140.563	-340.121
Wald χ^2	92.19	198.14
$p>\chi^2$	0.0000	0.0000

Note: *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

The probability of punishing other group members is predicted by greater physiological arousal and negatively-signed emotions. The estimation of a random-effects Probit model (model (1) in Table 4) shows that the more negative the emotions experienced when the individual receives information about his group members' contributions, the greater the probability of assigning punishment. Once this and other variables are controlled for, a higher SCR at the time of observing others' contributions also increases the probability of allocating punishment. Physiological arousal when making one's own contribution, however, does not significantly change this probability, whereas it increases significantly the severity of punishment assigned. This can be seen for model (2) in Table 4, in which the number of punishment points assigned to another group member is the dependent variable. Experiencing more positive emotions when contributing also significantly increases the severity of sanctions applied. One punishes more severely a group member who contributes below the average and whose contribution has generated a more negative emotional reaction.

3.5 The relationship between emotion and punishment for the sanctioner

Sanctioning others evokes emotions in the individuals who apply the sanctions. This can be seen in Figure 7. The figure shows that the more punishment points an individual assigns, the more he is physiologically aroused. The average SCR magnitude is 0.26 μS when no sanction is assigned, 0.31 μS when one is assigning one or two points, and 0.44 μS when one assigns more than two points in total to the other group members. A M-W test on period 1 individual data indicates that the SCR magnitude is significantly lower when no punishment points are assigned than when they are (p = 0.069), though the difference is not significant when the test is conducted at the group level (p = 0.335).

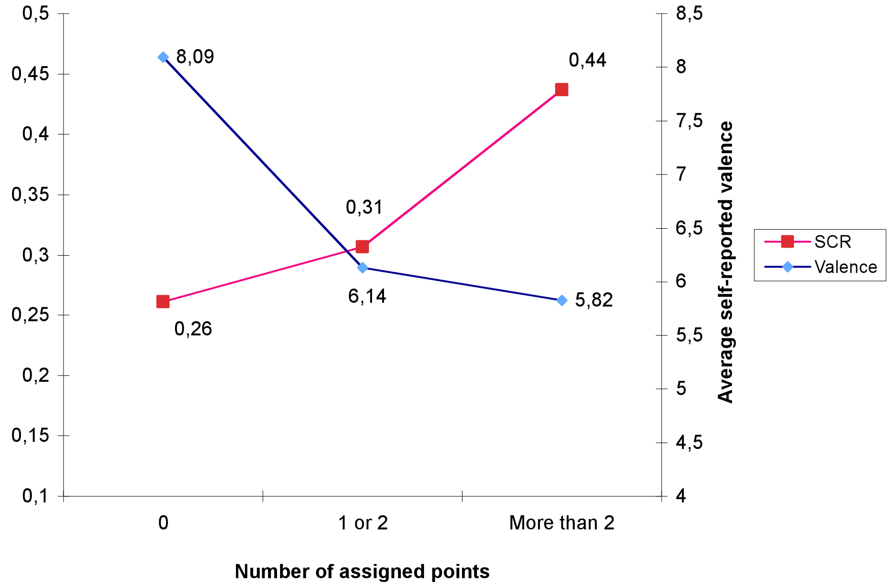

Fig.7. Physiological arousal (SCR) and self-reported hedonic valence of emotions as a function of the number of punishment points assigned to other group members.[4]

We also estimated a random-effects Tobit model, in which the dependent variable is the SCR magnitude when assigning punishment points. The estimation results are shown in Table 5. The independent variables include the SCR of the subject when he is informed of the contributions of other group members, the total number of punishment points assigned to group members, a dummy variable indicating whether points are assigned to people who contribute more than the punisher ('anti-social punishment'), a time trend, and the subject's gender (log-likelihood = -348.209, Wald χ^2 = 21.13, p < 0.001, N = 480, number of left-censored observations = 174).

[4] The average SCR magnitudes are displayed on the left vertical axis and the average self-reported emotional states are reported on the right vertical axis. Zero refers to extremely unpleasant feelings and 10 to extremely pleasant feelings. The number of points assigned to others is grouped into three categories: 0, 1-2 and greater than 2 points. As far as SCR measures are concerned, there are respectively 399, 54 and 27 observations in the categories of assigned points. When self-reports are considered, the numbers of observations are respectively 1327, 96 and 17. The discrepancy in the numbers of observations between SCR and self-reports is due to the fact that self-reports have been collected for each individual punishment assignment, whereas only one observation on SCR per player is obtained in the sanctioning stage.

Table 5. Determinants of the physiological arousal and the valence of emotions when assigning sanctions

Dependent variables	SCR (1)	Valence (2)	Valence (3)
SCR when receiving feedback about others' contributions	0.068 (0.055)	-	-
Valence of emotions when receiving feedback about others' contributions	-	-	0.335*** (0.056)
Punishment points assigned to group members	0.072*** (0.023)	-	-
Punishment points assigned to cooperators	-0.111 (0.115)	-	-
Punishment points assigned to *j*	-	- 0.371** (0.150)	0.339** (0.134)
Punishment points assigned to *j* if *j* is a cooperator	-	-0.163 (0.372)	-0.806* (0.469)
Time trend (*t*)	-0.003 (0.005)	0.015 (0.017)	0.006 (0.009)
Male	0.421*** (0.156)	-0.796*** (0.286)	-0.928** (0.395)
Constant	-0.065 (0.107)	-	-
N	480	1440	1440
Left-censored observations	174	-	-
Log-likelihood (pseudo)	-348.209	-2338.361	-2181.851
R^2 (pseudo)	-	0.044	0.108
Wald χ^2	21.13	15.92	189.92
$p>\chi^2$	0.000	0.003	0.000

Note: Model (1) is a random-effects Tobit model. Models (2) and (3) are ordered Probit models with robust standard errors (in parentheses) and clustering at the group level. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

The estimates indicate that the greater the number of points assigned, the greater the physiological arousal (coeff. = 0.072, $p = 0.001$). There is no additional effect of punishing a cooperator ($p = 0.335$). The conclusions are similar if a cooperator is defined as someone who contributes more than the average of the group. There is no significant trend over time ($p = 0.570$) and there is no correlation between the physiological arousal when assigning points and when learning of others' contributions ($p = 0.213$).

Figure 7 also shows that assigning punishment points is associated with less pleasant self-reported emotions. Indeed, the average valence is 8.09 for individuals who assign no sanctions, 6.14 for those assigning one or two points, and 5.82 for those allotting more than two points to another group member (a M-W test indicates that valence is significantly greater when no sanction is assigned, $p = 0.016$ when taking each group as an observation, and $p = 0.006$ with each subject in period 1 as an observation, two-sided).

At first sight, these results seem to contradict the findings by de Quervain *et al.* (2004) who have identified a "sweet taste of revenge". However, these findings and ours can be reconciled. First, the average valence of emotions increases significantly from 4.964 to 6.088 when subjects assign punishment points, compared with the moment when they learn others' contributions (M-W, $p = 0.078$). This indicates that punishing others increases the satisfaction of the subjects who have been hurt by learning the contribution of others. In contrast, the subjects who do not sanction experience deterioration of their emotional state. They report an average hedonic valence of 9.333 at the time of being informed of others' contributions and 8.094 at the time of the sanction decision (M-W, $p = 0.004$).

Second, to understand whether these less positive feelings when punishing are due to the very act of punishing or to the negative feelings associated with the low contributions of the targets, two ordered Probit models have been estimated. The models specify robust standard errors and clustering at the group level. In model (2), shown in Table 6, the independent variables include the number of punishment points assigned to another individual, and a dummy variable indicating anti-social punishment, which occurs when the target contributes more than the punisher. A time trend and the sanctioner's gender are also included. Model (3), in addition to the above variables, also includes the self-reported valence when the sanctioner is informed of the targeted player's contribution.

Without controlling for the valence of emotions when being informed of another individual's contribution, model (2) shows that a more severe punishment is associated with more negative feelings. This pattern is consistent with figure 8. However, the opposite is found in model (3), when the valence of the emotions experienced when being informed of the contribution of the group member is controlled for. This pattern can be easily explained. A negative valence of feelings, experienced when learning the contribution of an individual, contributes to a negative valence of emotions at the time of deciding on how much to punish the individual (coeff. = 0.335, $p<0.001$). Controlling for this, the more punishment points assigned, the more positive the reported valence (coeff. = 0.339, $p = 0.012$). This suggests that the valence of emotions experienced when sanctioning is influenced both by the negative feelings associated with the free-riding of a group member and by the satisfaction of punishing the free-rider. These results are therefore consistent with previous findings as they indicate that punishing reduces the disutility experienced when learning about others' free-riding (de Quervain et al., 2004). Anti-social punishment of cooperators, however, is accompanied by relatively negative emotions (coeff. = -0.806, $p = 0.086$, N=6).

3.6 Emotional and behavioral impact of the receipt of sanctions

Both the physiological arousal and the valence of experienced emotions indicate that receiving punishment triggers negative emotions. This is shown in figure 8. The average SCR is 0.142 (S.D. = 0.365) when a participant is not punished and equals 0.205 (S.D. = 0.389) when he is punished. The more strongly an individual is punished, the greater is his SCR.

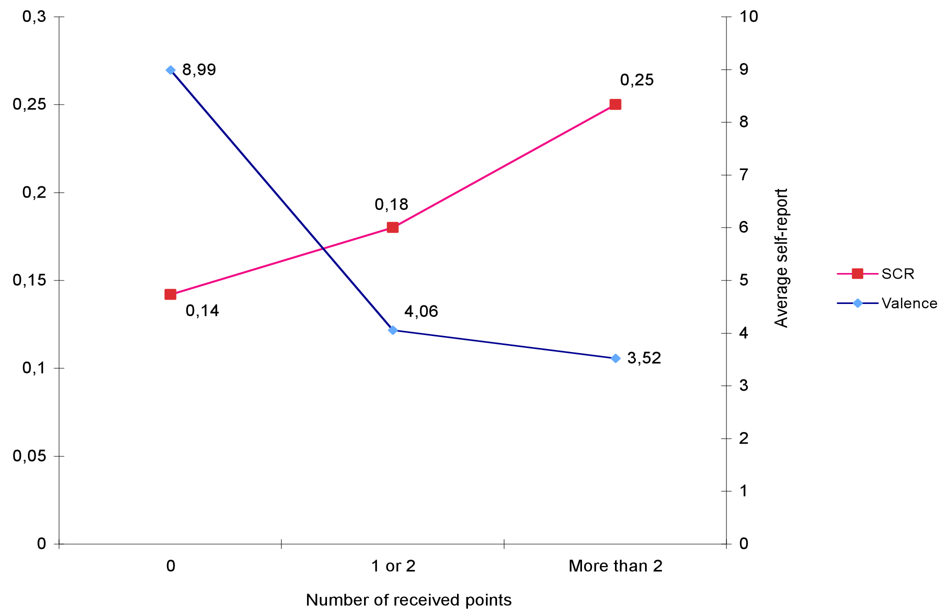

Fig.8. Physiological arousal and self-reported valence as a function of the number of punishment points received from other group members.[5]

A regression analysis of the SCR magnitude is presented in Table A in Appendix C (random-effects Tobit model, log-likelihood = -331.579, N = 480). It shows that the physiological arousal is positively related to the number of punishment points received (coeff. = 0.065, p = 0.051). The subjects who have been punished while contributing more than the average of the other group members experience the same physiological reactions as other individuals who are punished (p = 0.260). The SCR level declines over time, possibly related to the convergence to full cooperation (coeff. = -0.022, p = 0.003).

The valence of emotions, when informed about the sanctions one is receiving, is 8.99 when the individual is not punished (S.D. = 1.541) and 3.89 when he is punished (S.D. = 2.405). The difference is highly significant (M-W, p = 0.004). The valence patterns are further characterized with an ordered Probit model (see model (2) of Table A in Appendix C; pseudo-R^2 = 0.100, pseudo-likelihood = -711.863, N = 1440). The number of punishment points a person receives exerts a strong negative impact on the valence of her emotions (coeff. = -0.520, p<0.001). This effect is further reinforced if the punishment is anti-social, that is, if the recipient has contributed more than the group average (coeff. = -1.100, p<0.001).

[5] The points received are grouped into three categories, where 0 indicates that the participant has not been punished. Each point received decreases the recipient's payoff by ten percent. There are respectively 410, 47 and 23 observations in the three categories. The average SCR magnitudes are displayed on the left vertical axis. The average self-reported emotional valence is reported on the right vertical axis, where 0 corresponds to extremely unpleasant feelings and 10 to extremely pleasant feelings.

As already observed in previous studies (Fehr and Gaechter, 2000; Masclet et al., 2003), the change in individual contributions from one period to the next depends on the number of punishment points the individual received in the previous period. Indeed, a participant who was not sanctioned in the previous period, decreases his contribution in the current period by on average -0.12 ECU (N = 388; S.D. = 1.470); a participant who received between one and two points increases his contribution on average by 1.11 ECU (N = 47; S.D. = 3.737); and a participant who received three points or more increases it by 5.76 (N = 21; S.D. = 8.354).

The emotions experienced during various stages of the game influence subsequent cooperation levels. We estimated four models of the change in individuals' contributions between periods *t*-1 and *t* (models (1) are for the Baseline treatment and models (2) are for the Sanction treatment, see Table 6). The estimations are conducted separately for the individuals who contributed an amount greater than or equal to the group average (whom we will refer to as high contributors) and those who contributed less than the average (low contributors). The explanatory variables include the SCR magnitude and the valence of emotions when contributing in the previous period. In the Sanction treatment, the regressions also take into account the number of points received, as well as the SCR magnitude and the associated valence of emotions when receiving sanctions in the previous period.

Table 6. Emotions, punishment, and the change in contribution between periods t-1 and t

Dependent variable: Change in contribution between *t*-1 and *t*	Baseline treatment (1)		Sanction treatment (2)	
	High contributors	Low contributors	High contributors	Low contributors
Difference between own contribution and the average contribution in *t*-1, $(\overline{c}_{-i} - c_i)_{t-1}$	-0.750*** (0.088)	-0.523*** (0.149)	-0.065 (0.050)	-0.800*** (0.056)
Physiological arousal when contributing in *t*-1	-0.535 (0.611)	0.738 (1.450)	-0.216 (0.228)	0.204 (1.348)
Valence of emotions when contributing in *t*-1	0.248*** (0.077)	-0.204 (0.238)	0.013 (0.033)	0.199 (0.227)
Punishment points received in *t*-1	-	-	-1.258*** (0.397)	0.377*** (0.107)
Physiological arousal at the time of receipt of sanctions in *t*-1	-	-	0.266 (0.270)	3.391** (1.677)
Valence of emotions at the time of receipt of sanctions in *t*-1	-	-	-0.170** (0.078)	0.086 (0.288)
Time trend (*t*)	0.007 (0.056)	-0.041 (0.068)	-0.029 (0.021)	-0.013 (0.027)
Male	-1.028 (1.025)	-0.767 (1.632)	-0.458** (0.212)	0.058 (0.859)
Constant	-0.899 (1.085)	2.524 (1.652)	1.882* (0.979)	-3.040*** (1.158)
N	260	196	411	45

R^2	0.237	0.104	0.103	0.729
Wald χ^2	9194.90	144.14	183.03	67.65
$p>\chi^2$	0.0000	0.0000	0.0000	0.0000

Note: All the models are random-effects GLS models with robust standard errors (in parentheses) and clustering at the group level. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

In the Baseline, but not in the Sanction treatment, the more positive the emotions when contributing in the previous period, the more the high contributors increase (or rather the less they decrease since on average the change is negative) their contribution in the current period. The sanctions received, the SCR magnitude, and the valence of emotions when being informed of the quantity of sanctions received are strong predictors of changes in behavior in the Sanction treatment. Receiving sanctions motivates the low contributors to increase their contribution, whereas the high contributors adjust it downward in response. Controlling for this direct effect of sanctions, the more physiologically aroused the low contributors were when informed about how much they have been punished, the greater the upward adjustment of contribution in the next period. The valence of emotions exerts no significant influence. While SCR level does not predict the high contributors' decisions, the less negative their self-reported feelings when being informed about the sanctions they receive, the lower the upward adjustment of their contribution. Individuals dislike being punished and they adjust their behavior accordingly.

4. DISCUSSION AND CONCLUSION

We consider the emotional patterns associated with behavior in the Voluntary Contributions Mechanism, a well-known social dilemma. In our experiment, we record two measures of emotion, skin conductance response and self-reported valence of emotional state, as our subjects choose how much to cooperate, observe others' levels of cooperation, sanction others, and are themselves sanctioned. Our results show that higher contributions are associated with more pleasant self-reports on the part of the contributor. For a given contribution level, individuals are more satisfied in a setting, in which sanctions are permitted. Sanctioning non-cooperators mitigates and partially relieves the negative emotions associated with others' free-riding. Relatively negative emotions when receiving sanctions are associated with more cooperative future play. An individual's arousal level is greater, the less she contributes, the more others contribute relative to her in the absence of sanctions, and the more her contribution differs from others' when sanctions are possible. Stronger punishment assignments are associated with greater arousal both before and during the punishment decision. Greater punishment received leads to higher arousal, which in turn leads to a greater increase in contributions in the subsequent period. Emotions, therefore, appear to play a role in the emergence and maintenance of cooperation.

However, while emotions may nudge groups in the direction of cooperation, appropriate emotional responses of members of a group are not sufficient on their own to allow the group to attain a cooperative outcome. It appears that appropriate institutions must be in place to allow emotions to express themselves in behavior that can be communicated to other parties (Ostrom et al., 1990). Emotions cannot make cooperation emerge when no sanctions or other forms of communication are available, as the Baseline treatment shows. In the Sanction treatment, the mechanism whereby emotions induce cooperation operates in conjunction with the ability to punish free-riders. Emotional responses to free-riding appear to induce individuals to make use of punishment opportunities when they are available. The application of punishment can then set in motion a "virtuous emotional circle" that promotes cooperation. Cooperators in groups with free-riders, that is, those with high arousal and negative feelings, relieve negative emotions when they punish free riders. The free-riders experience negative emotions in response to punishment, and increase their subsequent level of cooperation. The new, higher, average level of contribution becomes the new standard or norm. Contributions below this level induce negative emotions in others and attract sanctions. Thus, both positive and negative emotional reinforcement, in conjunction with the availability of sanctions, encourages cooperation. Cooperation is associated with positive emotions for oneself and other affected parties, and receiving punishment is associated with negative emotions.

The behavior in the game that leads to positive valence of emotions is summarized in Table 7. In the table, the dependent variable is the overall valence of emotions averaged at the various moments of interest (contribution decision, receipt of information about others' contributions, decision to assign sanctions to other group members and receipt of information about sanctions received) within a period. This average valence is a measure of the overall satisfaction of individuals with the activity in the current period. In both treatments, the more people cooperate, the higher the average valence of emotions. The valence decreases when individuals contribute more than the other members of their group. It increases when individuals assign sanctions, especially when the group average contribution is relatively low. In the Baseline treatment, where no punishment institution is able to support cooperation within groups of individuals, the valence of emotions decreases over time. In the Sanction treatment, the time trend is positive.

The results here complement several of the results reported in earlier research on the physiological and emotional patterns that accompany cooperation and punishment. The willingness of people to voluntarily incur costs to punish opportunists has been related to the activation of the dorsal striatum in the brain in a positron emission tomography study (de Quervain *et al.*, 2004). This has been interpreted as evidence of satisfaction from punishing non-cooperators. We find evidence of similar satisfaction from punishing non-cooperators in our VCM interaction. But we also show that the satisfaction derived from punishment is not sufficient to compensate totally the negative valuation of others' free riding. Shame and guilt on the part of the target of sanctions have been shown to contribute to make punishment a deterrent of uncooperative behavior (Hopfensitz and Reuben, 2009). We also find that receiving punishment is associated with an adverse emotional state, and those in the most negative state, as registered with our measure, have the strongest tendency to modify their subsequent behavior in the direction of greater cooperation.

Table 7. Determinants of the average valence of emotions

Dependent variable: Average individual valence in a period	Baseline treatment (1)	Sanction treatment (2)
Mean contribution of other group members $\overline{c}_{-i}$	0.291*** (0.022)	0.303*** (0.039)
Difference between own contribution and the mean contribution of others (c_i - $\overline{c}_{-i}$)	- 0.026* (0.014)	- 0.075** (0.031)
Punishment points received	-	-0.251*** (0.035)
Total number of punishment points assigned	-	0.537***(0.195)
Total number of punishment points assigned * Mean contribution of group members	-	-0.038*** (0.012)
Time trend (t)	-0.036*** (0.010)	0.005 (0.020)
Male	0.238 (0.387)	-0.820*** (0.122)
Constant	2.847*** (0.475)	2.877*** (0.819)
N	480	480
R^2	0.671	0.324
Rho	0.413	0.658

Note: Models (1) and (2) are random-effects GLS models with robust standard errors (in parentheses) and clustering at the group level. *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

We conclude with a thought experiment. Suppose that we interpret the valence of reported emotional responses to events as a measure of experienced utility. Then, we can evaluate whether the pattern of experienced utility is consistent with three different types of assumptions on individual preferences: homo œconomicus, social welfare maximization, and strong reciprocation (Bowles and Gintis, 2002; Fehr et al., 2004). For a homo œconomicus, the valence of emotional responses would be increasing in his own monetary payoffs. Therefore, the valence of emotional responses would correlate negatively with own contribution, positively with others' contributions, and negatively with both the assignment and the receipt of sanctions. The emotional responses of a social welfare maximizer would be solely and positively related to the total group payoff. Therefore, emotional responses would correlate positively with both the level of one's own contribution and of others' contributions, and negatively with both the assignment and the receipt of sanctions. Finally, the emotional response of a strong reciprocator would be positively related to other players' contributions and to one's own assignment of punishment of free-riders. The emotional response patterns observed in our experiment fit best to the profile associated with strong reciprocation, with the additional presumptions that all else equal, receiving sanctions is viewed negatively, and that, all else equal, individuals prefer to cooperate. This pattern suggests that reciprocation is not necessarily a strategy used instrumentally as a means to maximize own or group earnings. It is also supported with autonomic responses.

References

Anders S., M. Lotze, M. Erb, W. Grodd, N. Birbaumer (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human Brain Mapping* 23, 200-209.

Ben-Shakhar, G., G. Bornstein, A. Hopfensitz, and F. van Winden (2007). Reciprocity and emotions in bargaining using physiological and self-report measures. *Journal of Economic Psychology* 28(3), 314-323.

Blau, P.M., (1964). *Exchange and power in social life*, John Wiley and Sons, Inc., New York.

Bosman, R. and F. van Winden (2002) Emotional Hazard in a Power-to-take Experiment. *Economic Journal* 112, 147-169.

Boucsein, W. (1992). *Electrodermal Activity,* Plenum Series in Behavioral Psychology and Medicine, Plenum.

Bowles, S. and H. Gintis (2001). *The Economics of Shame and Punishment*. Mimeo, University of Massachusetts.

Bowles S. and H. Gintis (2004). The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Journal of Theoretical Biology* 65, 17-28.

Bradley, M.M., M. Codispoti, B.N. Cuthbert, and P.J. Lang, (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion* 1, 276-298.

Coleman, J.S. (1990). *Foundations of social theory*, Belknap, Cambridge, MA.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset Putnam Publishers, New York.

Dawson, M.E., A.M. Schell, and D.L. Filion (2000). The electrodermal system, in *Handbook of Psychophysiology*, J. T. Cacioppo, L.G. Tassinary, G.G. Berntson, Eds., 2nd ed. (Cambridge University Press, Cambridge.

De Quervain, D. J. F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr (2004). The neural basis of altruistic punishment. *Science* 305, 1254-1258.

Edelberg, R. (1967). Electrical properties of the skin, in *Methods in Psychophysiology*, C.C. Brown, Ed., Williams & Wilkens, Baltimore, 1-53.

Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature* 36(1), 47-74.

Fehr, E., U. Fischbacher, and S. Gaechter (2002) Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms. *Human Nature* 13, 1-25.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980-994.

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

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative?: evidence from a public goods experiment. *Economics Letters 71*(*3*), 397–404.

Fudenberg, D., and J. Tirole (1991). *Game Theory*. MIT Press, Cambridge, MA.

Gächter, S., E. Renner, and M. Sefton (2008). The Long-Run Benefits of Punishment. *Science* 322, 5 December, 1510.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments.in Forschung und wissenschaftliches Rechnen, K. Kremer, V. Macho, Eds., GWDG Bericht 63, Göttingen, Ges. für Wiss. Datenverarbeitung.

Hermann, B., S. Gächter, and C. Thöni (2008). Antisocial Punishment Across Societies. *Science* 319, 7 March, 1362-1367.Science 319, 1362 (2008).

Homans, G.C. (1961). *Social behavior: Its elementary forms*, Harcourt Brace & World, New York.

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

Hopfensitz, A. and E. Reuben (2009). The importance of emotions for the effectiveness of social punishment. *Economic Journal* 119, 1534-1559.

Keser, C., and F. Van Winden (2000). Conditional Cooperation and Voluntary Contributions to Public Goods, *Scandinavian Journal of Economics*, 102(1), 23-39.

Lang, P.J., M.M. Bradley, and B.N. Cuthbert (1992). A motivational analysis of emotion: reflex-cortex connections, *Psychological Science* 3, 44-49.

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). (2003) Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review* 93, 366-380.

Mehrabian, A., and J.A. Russell (1974). *Approach to Environmental Psychology*, MIT Press, Boston, MA.

Milinski, M., and B. Rockenbach (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444, 7 December, 718-723.

Milinski, M., and B. Rockenbach (2008) Human behaviour: Punisher pays. *Nature* 452, 20 March, 297-298.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge Univ. Press, Cambridge.

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.

Rand, D.G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak (2009). Positive Interactions Promote Public Cooperation. *Science* 325, 1272-1275.

Rilling, J.K, D.A. Gutman, T.R. Zeh, G. Pagnoni, G.S. Berns, and C.D. Kilts (2004). A neural basis for social cooperation. *Neuron* 35, 395-405.

Rilling, J.K, A.G. Sanfey, J.A. Aronson, L.E. Nystrom, and J.D. Cohen (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *NeuroReport* 15(15), 2539-2543.

Sanfey, A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom, and J.D. Cohen (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 5626, June 13, 1755-1758.

Yamagishi, T. (1986). The Provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology*, 51(1), 110-116.

Zeiliger, R. A Presentation of Regate, Internet Based Software for Experimental Economics, http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt, (GATE, Lyon, 2000).

Appendix A. Instructions to participants

The following are the instructions, translated from the original French, for the Sanction treatment. The instructions for the Baseline treatment are a subset of those printed here. The instructions describe the timing of decisions, the decision rules, and the process of payoff determination. Subjects had to answer several control questions at the end of the instructions to ensure that they understood the rules of the game.

Text of Instructions

Thank you for taking part in this experiment 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.

All of 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 the 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 total 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.

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.

Description of each period

Each period consists of two stages.

First stage. At the beginning of each period each participant receives an endowment of 20 ECU.

You, as well as the three other members of your group, simultaneously decide how much of your endowment you will contribute to the project, by indicating a number between 0 and 20, inclusive.

After all group members have made their decision, your screen will show you the total amount of ECU contributed to the project by the members of your group (including your own contribution). You are also informed of the amount contributed by each of the three other members of your group to the project. Note: the order in which each contribution is displayed is changed randomly in each period (in other words, the number that appears first on your screen does not always correspond to the decision of the same player). You are also informed about your first-stage payoff .

Your income consists of two parts:

- the amount of your endowment which you have kept for yourself (that is, 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 income in ECU in the first stage of each period is therefore:

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

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 case each member of the group receives an income for the first stage 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 an income of 40% (of 9 ECU) = 3.6 ECU from the project.

For each ECU that you keep for yourself you earn an income of 1 ECU. For every ECU you contribute to the project instead, the total contribution to the project increases by one ECU. Your income from the project will increase by 40% (of 1 ECU) = 0.4 ECU. The income of the other group members will also rise by 0.4 ECU each, so that the total income of the group from the project rises 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 other members to the project. For each ECU contributed by any member you earn 40% (1) = 0.4 ECU.

Second stage. You can, if you like, indicate your disapproval of members of your group by assigning points that reduce their first-stage payoff. 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 reduces her first-stage income by 10%. Similarly, your income can be modified if the other members of your group wish to do so.

You decide how many points to give to each of the other three group members to reduce their income or leave it unchanged. You must enter a value for each member, between 0 and 10 points. If you do not wish to change the income of a specific member, then you must enter 0.

If you distribute one point to a member, you reduce his first-stage payoff by 10%; if you assign two points, you reduce his payoff by 20%, etc. The number of points you assign determines by how much you are willing to reduce his first-stage payoff.

If you assign points, you pay a cost in ECU that depends on the number of points you distribute to each subject. The more points you give to any subject, the higher your costs. Your total costs are equal to the sum of the costs of distributing points to each of the other three group members. The following table illustrates the relationship between points distributed to a subject and the cost of doing so in ECU:

Number of punishment points	0	1	2	3	4	5	6	7	8	9	10
Cost to the punisher	0	1	2	4	6	9	12	16	20	25	30

Percentage of reduction of the target's payoff	0	10	20	30	40	50	60	70	80	90	100

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 25 ECU more; if you give the last subject no points, this does not cost you anything. In this example, the total cost of the assigned points is 27 ECU (2+25+0). These costs will be displayed on your screen.

Your final income in ECU in each period is calculated as follows:

If you received less than 10 points:
final income = (income from the 1st stage))*[(10 - number of received points)/10] – cost of points you assigned

If you received 10 or more points:
final income = - cost of points you assigned

The maximum number of points received that can count against you in a period is 10. For example, if you received 3 points from other group members your first-stage payoff will be reduced by 30%. If you received 4 points from other group members your first-stage payoff will be reduced by 40%. If you received 10 points or more, you will lose 100% of your first-stage payoff. In this case, you make a loss if you have assigned points to other members of your group. The amount of the loss equals the cost of the assigned points.

Therefore, your income in ECU at the end of the second stage can be negative, if the costs of the points you distribute exceed your income from the first stage. You can, however, avoid such losses with certainty through your own decisions.

Additional questions

Several times during each period, we will ask you to describe your feelings. You report your feelings on a scale scoring from 1 (extremely unpleasant feeling) to 10 (extremely pleasant feeling).

You are not paid based on the answers to these questions. They have no influence on the remainder of the session. We do, however, ask you to answer these questions sincerely and with care.

Summary: You receive an endowment. You decide how much you are willing to contribute to a project. You are then informed of the total amount the group contributed, the contribution of each other group member, and your income. You can assign points to express your disapproval. Several times during the period, you answer questions regarding your feelings at the time.

At the end of each period, the next period begins. You receive a new endowment of 20 ECU and you choose your contribution.

* * *

Thank you for answering the questionnaire on your understanding that has been distributed to you. 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.

Appendix B. Assignment of punishment points

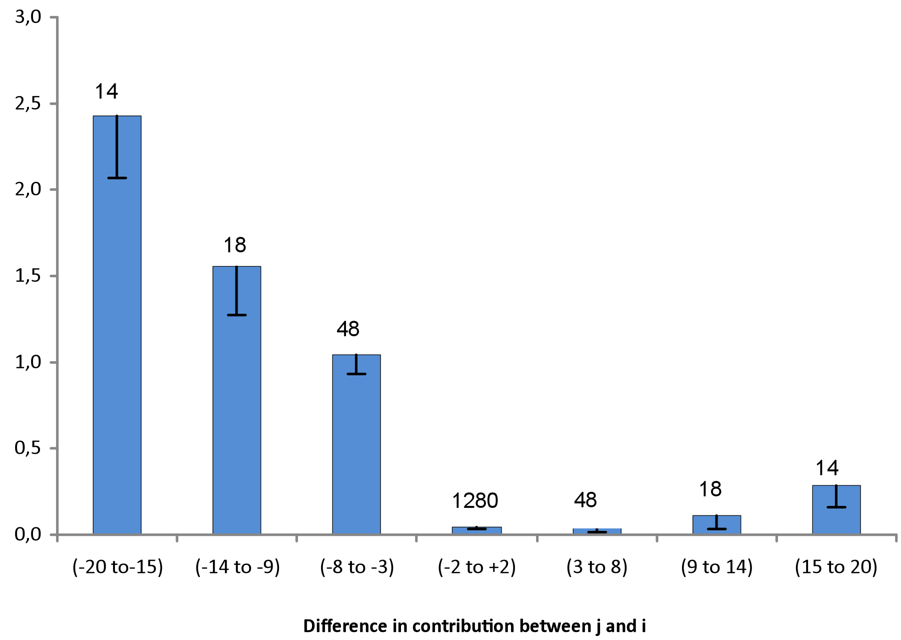

Fig.A. Number of punishment points assigned to j by i as a function of the difference in contribution between j and i

Figure A shows the average number of punishment points assigned by player *i* to player *j*. The observations are classified based on the difference between the contributions of players *j* and *i* in the current period. The horizontal axis is the difference in contributions, which, because each individual has 20 tokens, ranges from – 20 to + 20. The vertical axis indicates the average assignment of points, which can range from 0 to 10.

Appendix C – The receipt of sanctions

Table A. Determinants of the physiological arousal and the valence of emotions when receiving sanctions

Dependent variable	SCR (1)	Valence (2)
Punishment points received	0.065** (0.033)	-0.520*** (0.054)
Punishment points received by a subject contributing more than the group average	0.309 (0.275)	-1.100*** (0.315)
Time trend	-0.022*** (0.007)	0.015 (0.022)
Male	0.292* (0.156)	-0.107 (0.341)
Constant	-0.300** (0.125)	-
N	480	1440
Left-censored observations	320	-
Log-likelihood (pseudo)	-331.579	- 711.863
Pseudo R^2	-	0.100
Wald χ^2	22.00	265.01
$p>\chi^2$	0.000	0.000