1. Introduction

Reward and sanction mechanisms have been widely studied experimentally in social interactions environments. These mechanisms are seen as a prominent tool to try to achieve higher levels of cooperation among players. A huge literature has been focusing on financial sanctioning devices, introducing different costly punishment schemes, particularly in Public Good environments (see for example Fehr and Gaechter (2000)). If such mechanisms, when well designed, are efficient in their goal of raising contributions to the public good, they also entail monetary losses as punishment is destroying income both for the punished and the punishing players (Egas and Riedl (2008)). On the other hand, recent evidence suggests that the possibility of costly rewards is as efficient as a symmetric punishment scheme in raising contributions, but generates higher payoffs (Rand et al. (2009)).

As useful and interesting as this stream of research can be, one can question the availability of financial punishment during simple social interactions in real life. Monetary sanctioning may not always be available in such environments, as for example when a worker wants to express his disappointment toward a lazy colleague who is free-riding on his efforts. In order to tackle this realism issue but also to address the efficiency concern of costly punishment, a few papers have been interested in symbolic reward/sanction mechanisms. For example, Masclet et al. (2003) have made possible for players to send “disapproval points” to free-riders whereas Peeters and Vorsatz
(2009) implemented a system of “smileys” and “frownies”. If these devices are interesting for the understanding of non monetary feedback mechanisms, they are still very abstract and miss realism.

A very natural way to express disapproval or gratefulness after an interaction with somebody is to verbally communicate with him/her. Of course, if one wants to keep these interactions anonymous (as one often does in an experimental laboratory), face-to-face vocal communication cannot be implemented. An option is then to turn to written verbal feedback, e.g. written messages. Recent papers have investigated this by allowing players to send messages of their own composition in different one-shot games. Ellingsen and Johanessson (2008) showed that the possibility for receivers in a Dictator Game of sending a free message once the division has been known has an anticipated effect on dictators, raising donations (both by a decrease of “zero” donations and an increase of “fair splits”). Xiao and Houser (2005) found that receivers in a Ultimatum Game were accepting unfair offers more often when the possibility of sending a message was made available, suggesting a certain substitutability between financial and verbal punishment.

Our paper will try to address experimentally the ability of such a verbal feedback mechanism to maintain a certain level of cooperation in a Public Good Game, compared to a monetary punishment scheme. Moreover, we will investigate the role played by emotions in this context, both as an element which needs to be expressed and as a tool to try to trigger the desired behavior in the other.

2. Experimental Design

To tackle this question, we started from the design used by Fehr and Gaechter (2000). Each experimental session was divided in two parts (which order was counterbalanced between sessions). One part was devoted to the monetary Punishment Treatment (PT), as implemented by Fehr and Gaechter (2000), while the other was for the Communication Treatment (CT). In CT, instead of monetary punishment subjects could send a free form message to any of the four other members of their group. Each part was containing ten rounds. In the first phase of each round, subjects were asked the share of their endowment they wanted to contribute to the public good. After that, they received information about each member’s individual contribution. They could then, according to the treatment being played, either send costly points to any of their group members in order to destroy any proportion (from 0 to 100%) of his/her income or send a message of his choice to any of his group members.

Then, participants got to know if and how they have been punished or to read the messages they eventually received. After each message/punishment, subjects were asked to rate on a seven
points scale the intensity to which they experienced five different emotions (joy, sadness, pride, anger and shame/guilt) while they were reading a message or getting to know about the punishment points they eventually received. Once this was done, the next round could begin.

The composition of the groups was reshuffled after each round, making identification of a player (and thus reputation formation) impossible. This “stranger” matching procedure was also chosen so that communication could be used only as a feedback mechanism but not as a pre-play coordination tool for the next round. The subjects’ earnings were composed of a show up fee and of the payoff of a round from each part, selected randomly.

3. Preliminary Results

For the time being, three experimental sessions have been run, including 56 subjects. Another session, as well as control treatments (without any second phase after the public good game), will be run in a near future.

Contributions are significantly higher in PT (average: 58.7% of the initial endowment) than in CT (average: 31.2%), where approximately one third of all contributions are null. While the aggregate level of contributions stays stable during the ten rounds in PT, we observe a considerable drop in the second half of CT. Still at the aggregate level, we observe a characteristic punishing behavior in PT, with the use of punishment points decreasing over rounds (from more than one point sent per subject and per round in the first round to less than 0.25 in the last round). This is due to a drop in both the number of players being punished and the size of the punishment along the course of PT. On the other hand, it is hard to find a time trend in the use of messages in CT but this possibility has been quite widely used with more than one message sent on average (per subject and per round). As far as efficiency is concerned, payoffs in CT are inferior by 12% to those in PT and this difference does not appear statistically significant (but it may be explained by the lack of independent observations).

But what are the determinants of individual behavior in the feedback phase? Quite intuitively, a subject will send (resp. receive) more messages or punishment points when he contributed more (resp. less) than the average of his/her group. The emotional consequences of these messages or points are not the same depending on the treatment. Messages or points sent by a cooperative player (e.g. who contributed above group average) triggers in the recipient less anger both in PT and CT, reflecting the fact that such a player has “good reasons” to complain. But we also observe that only messages (not punishment points) sent by a cooperative player provoke more shame or guilt in the recipient: it seems that a punished free rider feel that he paid his dues and do
not experience any guilt while he does feel more guilty if he only received a message with no financial consequences.

Are these mechanisms efficient in affecting the dynamics of individual contribution choices? We indeed see that the punishment points received affects positively the change in contribution between two rounds. This is also the case for the number of messages received in CT: the more messages a player will receive, the more he will increase is contribution in the next round. Interestingly, two emotions seem to have contradictory effects in this respect: while the shame or guilt triggered by received messages have a positive effect on the change in contribution, anger provoking messages appear to be counterproductive. Importantly, this positive effect of messages is only present in the first half of CT: after some rounds, targets of messages do not change their behavior according to the number of messages they received anymore, probably reflecting a habituation to receive complaints about their behavior. Contrarily, punishment points stay effective over the ten rounds.

4. Discussion

The possibility to express emotions through messages does not appear as an efficient device to obtain high levels of contributions in a stranger matching public good game environment, especially compared to a monetary punishment mechanism. Although messages have a positive impact on the dynamic change in contributions (based on the shame they trigger in the recipient, while anger has a counterproductive effect), this effect is not permanent and vanishes in the second half of the experiment. Interestingly however, the two treatments are performing quite similarly in terms of efficiency.

Future sessions will show us if such a device can perform better than a simple public good game (without any communication or punishment). Next steps in this research direction include a system where both tools are made available to subjects: the dynamics in the use of these two alternatives as well as the effect on efficiency appear as very interesting questions.

5. References


