Journal of Economic Behavior and Organization 217 (2024) 631–646

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

The role of emotions in public goods games with and without punishment opportunities[☆]

Charles N. Noussair[a], Steven Tucker[b,*], Yilong Xu[c], Adriana Breaban[d]

[a] Eller College of Management, University of Arizona, 1130 E. Helen St. | P.O. Box 210108. Tucson, AZ 85721-0108, United States
[b] Waikato Management School, University of Waikato, Private Bag 3105. Hamilton 3240, New Zealand
[c] Utrecht University School of Economics, Kriekenpitplein 21-22, 3584 EC Utrecht, The Netherlands
[d] Erasmus School of Economics, Erasmus University Rotterdam, Oudlaan 50 3062 PA Rotterdam, The Netherlands

ARTICLE INFO

JEL codes:
D91
H41
C91

Keywords:
Public goods
Sanctions
Emotions

ABSTRACT

We consider the emotional correlates of activity in the Public Good game when monetary and non-monetary punishments are available, and when no punishment is possible. In our experiment, emotions are measured using Face Reading software that tracks the emotional content of facial expressions in real time. When no punishment is possible, greater anger and more negative emotional valence correlate with learning that one has contributed more than others. Lower valence and happiness, in turn, are associated with reducing one's cooperation in the next period. When non-monetary punishment in the form of expressed disapproval is possible, positive emotional valence is associated with cooperation, punishment of free-riders, and an increase in cooperation from one period to the next. Negative valence, on the other hand, is associated with the receipt of punishment, suggesting that the expression of disapproval inherent in the non-monetary punishment was well understood and had an effect on the emotions of the recipient. The data support the conjecture that the reinforcement that positive emotion provides is what allows non-monetary punishment to increase cooperation. In contrast, when monetary punishment is available, emotional correlates are less consistent, suggesting that monetary punishment is less reliant on emotions to be effective.

1. Introduction

The ability of peer punishment to promote cooperation has been one of the most celebrated findings in experimental economics. Beginning with the work of Yamagishi (1986), Ostrom et al. (1990) and Fehr and Gaechter (2000), numerous laboratory experiments have shown that allowing members of a group facing a social dilemma to reduce the earnings of individual other players, at their own expense, can promote cooperation within the group.[1] Individuals are very often willing to pay from their own earnings to punish free riders, and the existence of the possibility of punishment increases cooperation. These patterns of behavior clash sharply with the

☆ We gratefully acknowledge the financial support of the University of Waikato.
* Corresponding author.
E-mail address: stevent@waikato.ac.nz (S. Tucker).

[1] Endogenous choice of punishment institutions can increase the effectiveness of punishment, even in cultures where punishment is not so effective when these institutions are exogenously assigned (Gürdal et al., 2021). However, there are limits to this result. Punishment is not effective in increasing cooperation in many cultures (Herrmann et al., 2008), and if retribution by sanctioned parties is possible (Nikiforakis, 2008; Denant-Boemont et al., 2007).

https://doi.org/10.1016/j.jebo.2023.11.003
Received 1 June 2023; Received in revised form 25 September 2023; Accepted 2 November 2023
Available online 6 December 2023

subgame perfect Nash equilibrium, because the threat to punish is not credible. The decision to punish and its effect on cooperation can be rationalized by assuming that players have social preferences (Fehr and Schmidt, 1999) or gain utility from punishing others (Carpenter, 2007; Anderson and Putterman, 2006).[2]

A number of studies have also argued that specific emotional states accompany the decision to apply costly punishment against players who behave opportunistically (Bosman and van Winden, 2002; Casari and Luini, 2009; Dickinson and Masclet, 2015; Drouvelis and Grosskopf, 2016).[3] In particular, anger has been associated with a willingness to punish others (Xiao and Houser, 2005; Hopfensitz and Reuben, 2009; Nelissen and Zeelenberg, 2009). An emotion can be defined as "a positive or negative experience that is associated with a particular pattern of physiological activity" (Wikipedia entry for emotions). Perhaps unfortunately for economic analysis, emotions are a concept from psychology that does not translate well into the traditional economics vocabulary. In this regard, it is analogous to a word or phrase in a foreign language for which there is no exact translation into one's own language. Nevertheless, emotions are clearly concepts with meaning and constitute real, tangible forces affecting individuals' choices. A large body of literature in psychology and management documents that emotional state influences decisions in a wide variety of settings (see e.g. Lerner et al., 2015). The role of emotions in behavior is also increasingly being acknowledged in behavioral economics (Elster, 1998; Loewenstein et al., 2001; Battigalli and Dufwenberg, 2022). Emotional factors may be a crucial piece of the puzzle in understanding how it is that altruistic punishment leads to greater cooperation.

The effect of punishment on cooperation relies only partially on the fact that punishment reduces the pecuniary return to non-cooperation. Masclet et al. (2003), Noussair and Tucker (2005), and Mahmood et al. (2022) have shown that even symbolic sanctions that impose no pecuniary cost on either the sanctioning or sanctioned parties are consistently applied against free-riders and do have the effect of increasing contributions. The explanation offered by Masclet et al. (2003) for this effect is that symbolic sanctions communicate disapproval of the actions of the non-cooperator, and the non-cooperator is swayed by this disapproval (or the mere prospect of receiving disapproval) to increase their cooperation level. In our view, it is completely plausible that there are emotional correlates of the decision to contribute and to punish under this system, as well as emotional reactions to others' contributions and punishment one receives. However, to our knowledge, this has not been documented to date.

In this paper, we study the emotional state of participants as they play repeated trials of the public good game with monetary punishment studied by Fehr and Gaechter (2000), and the variation with non-monetary punishment of Masclet et al. (2003). We analyze the emotions that correlate with individuals' decisions: the levels of contributions and punishment assigned. We also study the emotional response to learning about others' contributions and the punishment that one has received. Additionally, we consider how emotional reactions in one trial of the game influence decisions in the next trial. We advance a number of hypotheses regarding how happiness and anger, the two emotions that have been most closely associated with cooperation and punishment, correlate with various events in the game. We also advance hypotheses about how emotional valence, a measure of overall net positivity of one's emotional state, correlates with activity in the game.

In contrast to previous work in this area, discussed in Section 2, we track the strength of specific emotions in real time over the course of the experiment. We also employ a new, non-invasive technology, Facereading, to acquire emotion data during our experiment. The Facereading software reads participants' facial expressions and classifies them based on the extent to which they conform to the profile associated with each of the six basic universal emotions: happiness, anger, fear, disgust, sadness, and surprise.

The use of Facereader software to elicit emotions has a number of potential benefits relative to conventional self-reporting methods. For one, we can measure and track emotional states at every instance of the experiment without interrupting the main activity. The possibility of not interrupting the public goods experiment with punishment is particularly crucial as the literature has suggested that a break, and thus a cooling off period, may reduce punishment (see, e.g., Dickinson and Masclet, 2015). Thus, we are able to capture instantaneous correlations between a (punishment/contribution) decision and an emotional state. Furthermore, self-reporting and describing one's own emotional state may itself affect emotions, as the explicit elicitation may induce experimenter demand effects (see, e.g., the discussion in Gross and Levenson, 1995) or be viewed as an unpleasant task that makes emotions more negative (Kugler et al., 2020). The accuracy of the software itself in uncovering emotional state has been validated in a number of studies (Bijlstra and Dotsch, 2011; Den Uyl and van Kuilenburg, 2005; Lewinski et al., 2014; Kuderna-Iulian et al., 2009; Terzis et al., 2010). Facereader output is highly correlated with psychophysiological measures of emotional valence (Facial Electromyography, EMG) and arousal (Skin Conductance Response, SCR) (Höfling et al., 2020).

The approach we take in this paper is to measure correlational, rather than causal, relationships. Much of the literature on emotions and cooperation has adopted a causal approach that involves inducing an emotional state at the outset of the session and then comparing decisions under different emotional states. This approach has the benefit of being able to identify the effect of an emotional state on behavior. For example, it allows the researcher to ask the question "If I make a group of individuals happier, does cooperation increase?". There are many important and interesting questions of this nature that the literature has investigated and will yet explore. Our approach of tracking emotional state and behavior in real time is well-suited to questions such as "are people happier as they are cooperating?" or "are people angrier when they are punishing". In our view, such correlational relationships, free of experimenter

[2] These two studies show that more punishment is assigned when it is cheaper for the sanctioner to reduce the earnings of the sanctioned party by a given amount. This indicates that the demand for punishment is downward-sloping at positive prices. This suggests that allocating punishment provides positive utility to the sanctioner at a decreasing rate, as is traditionally assumed for the utility of most goods under the classical theory of consumption.

[3] The notion that there are emotional correlates of behavior is fully compatible with the presence of social preferences which, like standard models of selfish preferences, are typically silent about emotional states.

intervention to induce emotional state, are also important in understanding the interplay between emotions, altruistic punishment, and cooperation, as well as the mechanism whereby altruistic punishment leads to higher cooperation.

In Section 2, we review the literature. In Section 3, we describe the experimental design. We present our hypotheses in Section 4. Section 5 contains the reporting of our results and Section 6 provides some concluding thoughts.

2. Literature review

The Public Good Game is a workhorse paradigm to measure the extent of cooperativeness in a group. The level of contribution is readily interpreted as indicating cooperativeness. In early work, Isaac and Walker (1988) and Andreoni (1988) showed that if the game is repeated, cooperation begins at an intermediate level and tends to decline over time. This decline has been attributed to initial errors and confusion (Ledyard, 1995; Houser and Kurzban, 2002; Andreoni, 1995) that decrease as individuals gain experience, and to the interaction of incomplete conditional cooperators and free riders (Fischbacher et al., 2001; Fischbacher and Gaechter, 2010).

To remedy this drift over time to greater free riding, peer punishment institutions have been proposed and studied (Yamagishi, 1986; Ostrom et al., 1990; Fehr and Gaechter, 2000; Fehr and Gächter, 2002). In such systems, group members observe the individual contributions of others and can pay from their own earnings to reduce the earnings of any other individuals. While punishment behavior and the reaction to it varies by culture (Hermann et al., 2008; Gürdal et al., 2021), in Western cultures there is a tendency to punish free riders and for the presence of a punishment option to increase cooperation.

Punishment exerts an effect both by reducing the pecuniary return to free riding and by communicating that the group disapproves of one's actions. Masclet et al. (2003) show that when punishment is assigned that imposes no monetary cost to either the sanctioning or sanctioned parties so that it serves purely as a means of expressing disapproval, cooperation also increases in the first 5 of the 10 periods in which the non-monetary punishment was in effect. Noussair and Tucker (2005), find that allowing both non-monetary and monetary punishment generates greater cooperation and earnings than either type of punishment alone. Mahmood et al. (2022) observe that the effect of non-monetary punishment weakens in teams, where a team refers to a pair of participants making a joint decision and each getting the same payoff. In other words, in their setting, two people play the role of one player in the game.

It has been argued that in some economic interactions, emotions may play a critical role in decision-making (Elster, 1998; Loewenstein, 2000; Thaler, 2000). A number of experimental studies have investigated the relationship between emotions and cooperation. Rilling et al. (2002), using a correlational approach, find that mutual cooperation in a Prisoner's Dilemma triggers activity in the dorsal striatum, an area of the brain strongly associated with reward, after the game is played. However, the predominant method to study the relationship between emotions and cooperation has been to induce an emotional state before the beginning of play. Groups are randomized into different emotional states to allow a causal effect of the emotion on cooperation to be observed. For example, de Hooge et al. (2007) find that induced guilt yields greater cooperation while shame has no effect. Nelissen and Zeelenberg (2009) probe the role of anger and guilt on third party sanction using two studies. They find that both induced anger (proxied by the intentionality of the norm-violation) and induced guilt (proxied by the responsibility for sanctioning) increase punishment. Punishment is reduced if either emotion is inhibited. Bartke et al. (2019) observe that inducing a state of caring increases cooperation, while Anger reduces it. Nguyen and Noussair (2022) find that induced happiness, disgust, and fear all lead to lower cooperation than a control condition.

A number of authors have proposed emotional underpinnings for the punishment of non-cooperators. Emotional states, particularly anger, correlate with the decision to punish, albeit in different paradigms than the one we study here. Xiao and Houser (2005) argue that costly punishment is used to express negative emotions in ultimatum games. Casari and Luini (2009) find that punishment is not used instrumentally, but rather purely as a response to the behavior of the punished individual, and they interpret the punishment behavior as governed by emotions. Hopfensitz and Rueben (2009) show that anger is a driver of punishment in a modified trust game. Galeotti (2015) finds that in a power-to-take game, which can be interpreted as an ultimatum game with continuous opportunities to punish, self- reported levels of anger, contempt and irritation partly explain punishment behavior. Sanfey et al. (2003) observe that recipients of an unfair offer in an ultimatum game experience increased brain activity in the anterior insula, which is associated with negatively-valenced emotional states. Additionally, the activation is a strong predictor of subsequent rejection of the offer.

Three studies have investigated the emotional correlates of activity in the game with monetary punishment that we study here. Drouvelis and Grosskopf (2016) induce anger and happiness before participants play the game once. They find that angry subjects contribute less than happy ones. They also observe that angry subjects punish more than those who are happy. Dickinson and Masclet (2015) study the game under four conditions. There is a Baseline condition which is the same as the Baseline in our study. In their Waiting treatment, designed to attenuate emotions, players must wait for 5 min after learning others' contributions before they punish. In the Waiting+Emotions treatment, in addition to waiting 5 min, players have the opportunity to vent their emotions via two avenues. The first is to privately indicate how angry, joyful and surprised they are at others' contributions, and the second is to write a message to the other members of their group, which they know will never be sent. In the Virtual Punishment treatment, they additionally have the opportunity to assign points of disapproval as in the Masclet et al. (2003) Non-Monetary Punishment treatment, though the points are not relayed to the recipient. They find that the more opportunities individuals have to express their emotions, the less monetary punishment is assigned. This result is interpreted as suggesting that monetary punishment is in part driven by the desire to express emotions. Joffily et al. (2014) use skin conductance responses and self-reported measures to track the arousal and emotional valence (overall positivity/negativity) of players over the course of the game. They find that the adverse emotional reaction that individuals experience when others free ride correlates with the assignment of sanctions, and the negative emotional reaction of those receiving sanctions correlates with their subsequently contributing more. They report that more positive emotions and lower arousal are correlated with higher contributions.

Our experiment differs from most of the above studies in the following ways. Rather than inducing emotions at the outset of play, we simply track them over the course of the experiment. Thus, instead of asking questions such as "does happiness increase cooperation?" we instead ask "what emotions are individuals experiencing as they cooperate?" Instead of only tracking arousal, in the form of skin conductance response, and overall valence of the emotional state, we additionally study specific individual emotions. The methodology we employ is Face Reading, which provides real-time tracking of the six basic universal emotions of all players, plus continuous measures of arousal and overall valence.[4]

3. Experimental design and laboratory procedures

Our experiment closely follows the design of Masclet et al. (2003), with a few parameter changes to simplify the experiment to better serve our purposes.[5] Throughout the entire experiment each participant is videotaped so that emotions in each stage of the experiment can be analyzed afterwards. In each experimental session, subjects are randomly assigned to groups of four and participate in a repeated voluntary contributions mechanism (VCM) for 10 periods. The composition of each group remains fixed for all periods.

At the beginning of each period, each agent is endowed with 20 Experimental Currency Units (hereafter, ECU), which can be converted into New Zealand Dollars at the end of the session using a predefined exchange rate. Subjects then simultaneously choose a portion of their endowment to contribute to a group account. Each ECU contributed to the group account yields a payoff of 0.4 ECU to each of the four members of the group. The remaining portion of the ECU that is not contributed to the group account goes to the subject's private account with no additional payoff to other group members. Hence, the earnings (π_t) of subject i in a period can be expressed as

$$\pi_t = 20 - c_i + 0.4 \times \sum_{k=1}^{4} c_k,$$

where c_i is the contribution of subject i. The total individual earnings in the experiment are the sum of the earnings over all periods. Clearly, individual i's earnings are maximized at $c_i = 0$. If the game is played once, the dominant strategy is to contribute zero. In a finitely repeated setting, the unique subgame-perfect equilibrium of the game is for all players to contribute zero in each period.

There are three treatments in the experiment. The Baseline treatment consists of the Public Good game described above played repeatedly for 10 periods. At the end of each period, subjects are shown a summary screen reporting their own contribution, the sum of all contributions, and their earnings for the period. In both the monetary punishment (MP) and the nonmonetary punishment (NP) treatments, the game continues onward to a punishment stage. Each subject is informed of the contribution levels of each of the other members of their group. We emphasize that the displayed order of the contribution decisions presented on subjects' screens is randomized in each period. This shuffling of positions is intended to force all punishment for behavior in a particular period to occur in the same period.

After reviewing the contribution decisions of others, subjects are given opportunities to express their views by assigning punishment points. In particular, in the monetary punishment (MP) treatment, subjects can assign up to 10 punishment points at a cost (1 point = 1 ECU) to each of the three other group members. Each punishment point received by a subject from any other group members reduces the earnings of the subject by 3 points. Subject i's earnings in a period in this treatment equal:

$$\pi_t = \left(20 - c_i + 0.4 \times \sum_{k=1}^{4} c_k\right) - \sum_{k \neq i} P_{ik} - 3 * \sum_{k \neq i} P_{ki}.$$

where P_{ik} is the punishment subject i assigns to subject k in her group, while P_{ki} is the punishment point subject i receives from subject k. The subgame-perfect equilibrium of the 10-period finitely repeated game is that all players always contribute zero and there is no punishment at all for any past history.

The Non-monetary Punishment (NP) treatment is identical to MP treatment, except that the punishment points assigned to a subject have no effect on her final earnings, and these punishment points are costless to assign. As in MP, subjects have the opportunity to assign between 0 and 10 points to each other group member. In both treatments, the points were presented as indicating a level of disapproval of a subject's contribution to the public account in the current period and were described to the subjects as such. The upper bound of 10 represents the highest level of disapproval and the lower bound of 0 points the lowest. Under nonmonetary punishment, all subgame-perfect equilibria for the finitely repeated game involve all players contributing zero regardless of past history, and any profile of disapproval point assignments, which are cheap talk, is consistent with an equilibrium.

[4] The six basic universal emotions, as proposed by Paul Ekman and coauthors, are Happiness, Anger, Fear, Sadness, Surprise, and Disgust. They are considered universal because a face expressing one of these emotions is recognized to be doing so in all human cultures, have analogues in other primates, and are common between blind and sighted people, indicating that they are innate rather than learned through social interaction (Ekman, 2007; Matsumoto and Willingham, 2009). Valence is a measure of the net positivity of emotional state. The Facereader software that we employ in our experiment calculates Valence as the value of the only positive emotion, Happiness (which is expressed on a scale from 0 to 1) minus the strongest of the four negative emotions (max{Fear, Disgust, Anger, Sadness}) each also expressed on a scale from 0 to 1. Thus, Valence can take any value on $[-1, 1]$ with greater values indicating a more positive emotional state. Arousal is a measure of autonomic activation. Facereader measures Arousal from a different profile of facial muscle movements than emotional states and expresses it on a scale from 0 to 1. Under the circumplex model (Russell, 1980), valence and arousal are completely orthogonal to each other.

[5] The instructions were identical to those used by Masclet et al. (2003).

Subjects' facial expressions were recorded and analyzed. This was done by videotaping the participants during their session and analyzing the videotapes later with Noldus FaceReader software. The software analyses the emotional state as revealed in participants' facial expressions. The time periods of interest are (i) before participants decide how much to contribute to the public account, (ii) after they learn what others have contributed, (iii) before they assign punishment, and (iv) after they learn the punishment that they have received.

The FaceReader software operates in the following manner. Every 1/10th of a second, an algorithm searches for a face in the videotape. When it detects a face, it applies a proprietary algorithm based on the Facial Action Coding System (FACS). The algorithm maps the vector of distances between 500+ points on the face to a vector in [0,1] which indicates a level of conformity of the facial expression to profiles associated with the six basic universal emotions: Happiness, Anger, Sadness, Fear, Disgust, and Surprise, as well as Neutrality[6] and Arousal. The weights do not necessarily sum to 1. It also calculates a measure of emotional Valence from the values of Happiness, the sole positive emotion, and Fear, Anger, Sadness, and Disgust, the four negative emotions. Valence takes on a value between −1 and 1, with higher values representing a more positive emotional state. FaceReader has been validated in a number of ways. The software has an accuracy rate of 90 percent when it rates the intended expressions of actors (Bijlstra and Dotsch, 2011). It correlates highly with self-reported natural unintentional emotions (Den Uyl and van Kuilenburg, 2005). It classifies human expressions as well as trained human observers (Kuderna-Iulian et al., 2009; Terzis et al., 2010; Lewinski et al., 2014). It has been used in a number of recent experimental economic studies (Fiala and Noussair, 2017; Van Leeuwen et al., 2018; Breaban and Noussair, 2018; Kugler et al., 2020).

Our experiment was conducted in the waikato experimental economics laboratory (WEEL) at the University of Waikato, New Zealand in the year 2016. A total of 112 subjects participated. There were two sessions conducted under each of the three treatments for a total of 6 sessions. Both NP and MP consisted of 10 groups and Baseline had 8 groups. Group membership was fixed for the entire session. Upon arrival, subjects were provided with a document called the Human Ethics Participation Information Sheet, a consent form and an instruction packet. The experimenter provided a verbal introduction, which included a brief welcome, discussion of the video recording process and a description of the procedures. We then read the instructions aloud and checked participants' understanding via several control questions.

At the beginning of each period, subjects received 20 ECU and were simultaneously asked to decide how much of the 20 ECU they would like to place in the public account. After all group members completed their contribution decision, subjects were presented with their individual contribution, sum of group contributions, and their own earnings. This sequence of activity constituted a period in the Baseline treatment. In the MP and NP treatments, subjects continued to a second stage where they could view how much other each group member contributed and had the opportunity to assign points to each other group member(s) if they wanted to. The period concluded with a summary screen presenting their own contribution, the sum of contributions in the group, the total amount of points they assigned, their cost of assigning points, the number of points they received, the reduction in their earnings from points received (this was always zero for NP), and their overall earnings for the period. Participants repeated the game for 10 periods. At the end of the session, subjects were called one by one to a private payment room to receive their earnings.

4. Hypotheses

The hypotheses for the experiment are based on interpretation or extrapolation of findings in the prior experimental literature. The work of Drouvelis and Grosskopf (2016) suggests that happiness causes more cooperation, while that of Nguyen and Noussair (2022) finds that happiness works in the opposite direction. Drouvelis and Grosskopf (2016), Bartke et al. (2019), and Nguyen and Noussair (2022) all find that anger leads to lower cooperation. Joffily et al. (2014) report that the valence of emotional state correlates positively with cooperation. Since the balance of the evidence associates positive emotions with more cooperation, we hypothesize that happiness would correlate with greater contributions. We also hypothesize that anger would correlate with lower contributions. As a consequence, emotional valence would correlate positively with contribution level.

Hypothesis 1. *An individual's Happiness and Valence level before the contribution decision are positively correlated with their contribution level. Anger before the contribution decision is negatively correlated with contribution level.*

In the game, the next change in the environment, after the contribution decision, occurs when individuals are informed of the contributions of others. Joffily et al. (2014) document that learning that one has contributed more relative to others triggers negative emotional valence, while contributing less than others leads to a more positive emotional state. We hypothesize that the same would be true for the specific emotions of the same valence.

Hypothesis 2. *The more that others in one's group contribute compared to one's own contribution, the greater one's Happiness and Valence. The less others contribute relative to one's own level, the greater one's Anger.*

Hypotheses 1 and 2 apply to all three treatments. For the MP and NP treatments, we also propose hypotheses regarding the emotional correlates of punishment behavior. Sanfey et al. (2003), Xiao and Houser (2005), Hopfensitz and Rueben (2009), and Drouvelis and Grosskopf (2016) find that the specific emotion of anger is linked to the application of punishment. We hypothesize that

[6] The Facereader provides measures of how closely a facial expression conforms to profiles associated with different emotional states. Neutrality is one of these profiles indicating the absence of the facial movements associated with the other six emotions. Conformity to Neutrality is reported on a scale from 0 to 1 by Facereader.

this relationship would also hold in our study with our technology for emotion measurement and generalize to a punishment that is non-monetary. We also predict that punishing improves the sanctioner's emotional state, as observed in the brain imaging study of de Quervain et al. (2004).

Hypothesis 3. *In both the MP and NP treatments, there is a positive correlation between the level of Anger an individual expresses after learning about the contribution of others and the punishment points that they assign. Furthermore, there is a positive correlation between (a) the changes in Happiness and Valence when the punishment is assigned, and (b) the number of punishment points that is allocated.*

Our final hypothesis concerns the emotional response to the receipt of punishment. Joffily et al. (2014) find that receiving punishment triggers greater arousal and a more negatively-valenced emotional state. We hypothesize that it would have the same effect in both our MP and NP treatments.

Hypothesis 4. *There is a positive correlation between the number of punishment points received and the change in Anger when the punishment is received. There is a negative correlation between the number of punishment points received and changes in Happiness and Valence.*

As a last topic of our analysis, we study whether there are emotional correlates of the decision to modify one's contributions between the current period and the next. The next section reports the results of our experiment.

5. Results

5.1. Contribution and punishment patterns

The average individual contribution levels by period in the three treatments are shown in Fig. 1. The figure reveals a number of patterns. In the Baseline treatment, the average contribution is close to 9 tokens out of a maximum possible of 20 in period 1. Contributions decline over time so that by period 10, they average approximately 3 tokens. This is highly consistent with previous results in the literature (Andreoni, 1988; Isaac and Walker, 1988). In the Monetary Punishment treatment, average cooperation also begins at an intermediate level of roughly 11 and increases over time to the 14 – 15 range. Monetary punishment is effective in generating cooperation, as has been documented in prior studies (Yamagishi, 1986; Ostrom et al., 1990; Fehr and Gaechter, 2000). The NP treatment also generates high levels of cooperation, but these begin to decline in period 3, and by the end of the sessions, contributions in NP lie between those in the MP and the Baseline treatments. This is also consistent with previous research (Masclet et al., 2003; Noussair and Tucker, 2005; Mahmood et al., 2022).

Fig. 2 shows the average level of punishment a person assigned to each of their group members in each period in the NP and MP treatments, $\sum_i \sum_j p_{ij}/n$. The figure reveals the following patterns. Under MP, the average punishment meted out is highest at the outset, at an average of 1.17 points assigned per individual and fluctuates over time around a level of approximately 1 point per period. Under NP, punishment points average about 9 at the outset of play, approximately 30 % of the maximum possible level of disapproval. Point assignments increase over time as it appears that this form of punishment sees its effect erode over time, and individuals express ever stronger levels of disapproval as contributions decline. Figs. 1 and 2, taken together, show that under both punishment processes, there is a strong negative relationship between the change in contributions in one period and the change in punishment in the same period.

5.2. Emotions and cooperation

We now turn to the emotional correlates of activity in the game. Hypothesis 1 asserted the presence of a number of relationships between emotional state prior to the contribution decision and the amount contributed. Table 1 reports the results of a regression

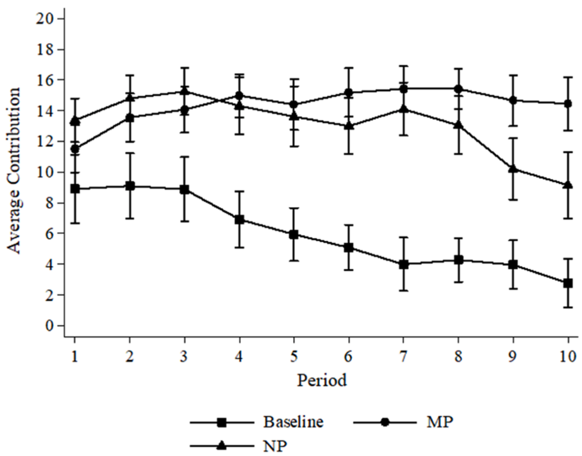

Fig. 1. Average Individual Contributions by Period in each Treatment.

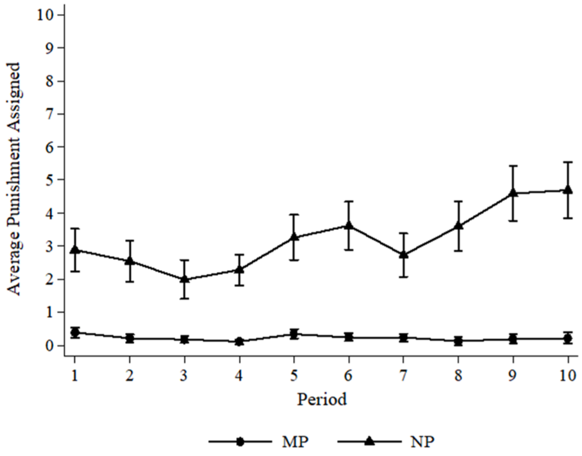

Fig. 2. Average Individual Punishment Assigned in Each Period of MP and NP.

analysis that considers how current emotional state correlates with contribution level, and which relationships are statistically significant. The following specification is estimated:

$$c_i^t = \beta_0 + \beta_1 Happy + \beta_2 Angry + \beta_3 Sad + \beta_4 Scared + \beta_5 Disgusted + \beta_6 Surprised + \beta_7 Arousal + \beta_8 Period$$

The estimates show that Happiness and Anger are not significant correlates of contribution in the Baseline treatment. In the MP treatment, Happiness is a significantly positive correlate, but Anger is not. The same pattern is observed under NP, with the coefficient of Happiness significantly positive and that of Anger insignificant. None of the other emotions are significant in any of the regressions. Thus, in two of the three conditions Happiness is associated with higher contributions. The coefficient of Happiness remains significant at $p < 0.05$ if one corrects for the testing of multiple (three) hypotheses. The last three specifications consider the correlation between Valence and contributions. The estimates show that Valence is not significantly correlated with contributions. The overall time trend, captured by the variable Period, is significantly negative under Baseline and NP, and significantly positive under MP. Thus, Hypothesis 1 receives mixed support. Happiness correlates with higher contribution, but only when punishment options are present. We detect no relationship between either Anger or overall Valence, and contributions. Therefore, Hypothesis 1 is supported for Happiness, but not for Anger or Valance. These findings constitute our first result:

Result 1: When punishment is available, Happiness is correlated with higher contributions, while Anger and Valence are uncorrelated with contributions. There is only weak support for Hypothesis 1.

We now turn to the emotional reaction to learning about the contribution of others. We focus on the difference between one's own contribution and the average of others', which has been shown to be strongly linked to the punishment that one assigns (Fehr and Gaechter, 2000; Masclet et al., 2003). We estimate the following specification:

$$\Delta E_i^t = \beta_0 + \beta_1 \left[c_i^t - \frac{\sum_j c_{j \neq i}^t}{n} \right],$$

where ΔE_i^t is the change in the emotion from the ten seconds preceding receipt of the information regarding other people's contribution to the ten seconds immediately subsequent to getting the information. Table 2 shows the estimates for the dependent variables of Happiness, Anger, and overall Valence of emotional state. The estimates from the Baseline treatment show that the difference between one's own and the group average contribution is significantly positively correlated with the anger that one subsequently feels. Anger, however, is not significant in either of the other two treatments. This lack of the significance of Anger in the punishment conditions may be a result of knowing that one can punish the free-riding behavior of others in the next stage, while one feels helpless to do so in the Baseline condition. Nevertheless, the effect is small, with Anger increasing by 0.02 standard deviations for each one unit change in the average contribution. Happiness is not responsive to information about others' contributions. Under Baseline, valence decreases when others contribute less relative to one's own contribution, as the higher level of anger reduces the value of the valence measure. Overall, Hypothesis 2 is not supported. This is our second main result.

Result 2: In the absence of the ability to punish, Anger is correlated with the extent to which one has contributed compared to the group average. Otherwise, Hypothesis 2 is not supported. The relationship does not appear when punishment is available. Happiness and valence do not show a correlation with one's contribution relative to the average.

We now consider whether *mutual* cooperation triggers an increase in Happiness or overall Valence, over and above any emotional response to the contribution of others alone. To evaluate this conjecture, we estimate the specification:

Table 1
Contribution as a Function of Emotions and Valence, By Treatment.

	Baseline	MP	NP	Baseline	MP	NP
Angry	−3.036	2.678	0.337			
	(3.199)	(2.441)	(3.447)			
Happy	0.996	9.931***	14.874***			
	(5.563)	(3.273)	(3.770)			
Sad	−0.463	2.507	−0.173			
	(2.082)	(1.683)	(3.575)			
Surprised	6.113	−4.962	−12.986			
	(11.754)	(9.069)	(14.984)			
Scared	−8.128	−12.587	−13.496			
	(11.196)	(13.319)	(16.907)			
Disgusted	1.933	1.806	−1.531			
	(4.112)	(1.986)	(8.140)			
Arousal	2.124	−2.940	−1.491			
	(3.714)	(3.009)	(2.763)			
Period	−0.750***	0.249***	−0.536***			
	(0.131)	(0.082)	(0.123)			
Constant	9.767***	12.501***	16.392***			
	(1.432)	(1.333)	(1.395)			
Valence				1.347	−0.196	2.256
				(2.126)	(1.422)	(2.724)
Period				−0.750***	0.261***	−0.499***
				(0.128)	(0.079)	(0.124)
Constant				10.463***	12.602***	16.398***
				(1.396)	(0.780)	(1.089)

Notes: Entries are estimated coefficients from random effect panel regressions (specification chosen with a Hausman test). The dependent variable is the level of contribution and the explanatory variables are the average emotions in the ten seconds before making a contribution. Standard errors, clustered at the individual level,[a] are reported in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[a] We believe that it is more natural to cluster at the individual level than at the group level when referring to the regressions concerning emotions. In our experimental environment with private decision making, emotions are also private. They are not observed by others. This means that one person's emotional state cannot influence another person's directly (they can only do so indirectly through their effect on behavior).

Table 2
Change in Emotional State as a Function of the Difference between One's Own Contribution and the Average Contribution of Others in the Group.

	Baseline	MP	NP	Baseline	MP	NP	Baseline	MP	NP
Anger	0.002***	−1.40E-04	−0.001						
	(0.0007)	(0.0007)	(0.0006)						
Happiness				−1.16E-04	−0.001	−2.56E-04			
				(0.0012)	(0.0006)	(0.0003)			
Valence							−0.002*	−0.001	−2.92E-04
							(0.0009)	(0.0020)	(0.0009)

Notes: Entries are estimated coefficients of the differences in contribution (own contribution minus the average contribution of others) from panel regressions. For the MP and NP treatments, a random effect model is employed. The Baseline data uses a fixed effect model, as suggested by a Hausman test. The dependent variable is the average change in emotion between the ten seconds before and ten seconds after observing other group members' contribution levels. Standard errors, clustered at individual level, are reported in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Entries with E-XX refer to 10^{-XX}, thus, 1.16E-04 is $1.16 * 10^{-4}$.

$$\Delta E_i^t = \beta_0 + \beta_1 c_i^t + \beta_2 \frac{\sum_j c_{j\neq i}^t}{n} + \beta_3 c_i^t * \frac{\sum_j c_{j\neq i}^t}{n}$$

The last term is the interaction of *i*'s own and the average contribution of others in her group. If mutual cooperation triggers happiness, the coefficient β_3 would be positive when the dependent variable is Happiness. If mutual cooperation diffuses anger, then coefficient β_3 would be negative when the dependent variable is Anger. As a consequence, β_3 would be positive for Valence. The estimates are shown in Table 3.

The estimates show that the interaction term is not significant in any treatment for Happiness, Anger or Valence. Thus, one's emotional response to others' cooperation is similar whether one has been cooperative or non-cooperative oneself. We obtain no evidence that there is a specific emotional response to mutual cooperation. Under MP, higher contributions on the part of others

Table 3
The Effect Contribution Decisions on the Change in Emotional State.

	Baseline			MP			NP		
	Anger	Happiness	Valence	Anger	Happiness	Valence	Anger	Happiness	Valence
Contribution	0.002	−2.74E-04	−0.004	−0.002	−0.001	0.004	−0.003	−0.001	−0.001
	(0.0021)	(0.0020)	(0.0029)	(0.0033)	(0.0017)	(0.0052)	(0.0025)	(0.0010)	(0.0030)
Other's Contribution	−0.002	−0.001	−0.003	−0.004	0.001	0.009**	−0.003	−0.001	1.64E-06
	(0.0015)	(0.0011)	(0.0024)	(0.0035)	(0.0014)	(0.0046)	(0.0027)	(0.0010)	(0.0035)
Interaction	1.19E-04	−7.36E-05	−7.27E-05	1.05E-04	6.77E-05	−3.30E-04	1.66E-04	5.08E-05	9.98E-05
	(0.0003)	(0.0002)	(0.0004)	(0.0002)	(0.0001)	(0.0003)	(0.0002)	(0.0001)	(0.0002)
Constant	0.011	0.025	0.078***	0.061	−0.007	−0.108*	0.060	0.015	0.012
	(0.0141)	(0.0122)	(0.0168)	(0.0156)	(0.0151)	(0.0580)	(0.0404)	(0.0130)	(0.0468)

Notes: Entries are estimated coefficients from panel regressions. For the MP and NP treatments, a random effect model is employed. Two regressions on the baseline data used a fixed effect model (specifically for the changes of Happiness and Anger), as suggested by a Hausman test. The dependent variable is the average change in an emotion between ten seconds before and ten seconds after seeing other group members' contribution levels. Standard errors are clustered at the individual level. Standard errors are reported in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Entries with E-XX denote 10^{-XX}, thus, 1.19E-04 is $1.19 * 10^{-4}$.

significantly increases Valence.[7]

5.3. Emotions and punishment

We now turn to the emotional correlates of the application of punishment. Fig. 3 compares the level of Happiness, Anger, and Valence of those who did not punish and those who assigned various levels of punishment.[8] The figure suggests that greater Happiness and overall Valence[9] are associated with lower point assignments. Greater Anger accompanies greater punishment in NP.

We estimate the following regression specification to consider whether these relationships are statistically significant and to identify any other emotions that are correlated with the assignment of punishment.

$$P_{ij}^{t} = \beta_0 + \beta_1 Happy + \beta_2 Angry + \beta_3 Sad + \beta_4 Scared + \beta_5 Disgusted + \beta_6 Surprised + \beta_7 Arousal + \beta_8 Period$$

where P_{ij}^{t} is the number of punishment points that player i assigns to j in period t. In some specifications, the independent variable $c_j^t - \frac{\sum_j c_{j \neq i}^t}{n}$ (Difference in Contribution) is added, since the difference between one's own contribution and the group average is known to be a strong predictor of the punishment that one assigns (Fehr and Gaechter, 2000; Masclet et al., 2003). The emotions are averaged over the ten seconds immediately preceding the assignment of points to other players. The results are presented in Table 4.

The estimates show that under both treatments, Valence is negatively associated with punishment. Individuals in a more negative emotional state punish more. While Anger and Happiness are not significant contributors to this effect, Sadness, Fear and Disgust consistently exhibit positive relationships, significantly so under some specifications, with the amount of punishment assigned. It appears to be that it is these negative emotions that account for the relationship between Valence and punishment. As illustrated in Fig. 3, punishments increase over time under NP. The variable Difference in Contribution is positive and significant, indicating a tendency for individuals to direct punishment to those who have contributed less than they have.

As mentioned earlier, it has been proposed that there can be an improvement in emotional state after one has assigned sanctions to other parties (de Quervain et al., 2004). We consider this in our data. Fig. 4 shows the change in Happiness, Anger and Valence, between the ten seconds preceding and the ten seconds subsequent to the punishment decision, for different levels of punishment. The figure gives the impression in both treatments, that the highest punishers have larger decreases in Anger and increases in Valence than other players.

Regressions of the change in emotional state (between ten seconds after and ten seconds before assigning the punishment points) as a function of the number of points assigned are shown in Table 5. We find that there is a positive correlation between the number of punishment points assigned and the change in Valence in both MP and NP. The relationships between Happiness and Anger with points assigned are not significant. Thus, while Hypothesis 3 is supported for Valence, it is not upheld for Anger and Happiness.

[7] Analysis of the changes in each of the other specific emotions involved reveals that the increase in Valence is primarily taking the form of a decrease in Sadness.

[8] The different bins into which punishment levels are divided were chosen based on the number of observations, since the distribution of assigned points is skewed towards low values.

[9] The negative valence overall in the experimental session is a typical pattern observed throughout different studies that use FaceReader software, such as in auction experiments (Breaban et al., 2022), asset market experiments (Breaban and Noussair, 2018) or ultimatum games (van Leeuwen et al., 2018). Experiments appear to be an activity requiring concentration and attention from participants that typically leaves them in a negative emotional state.

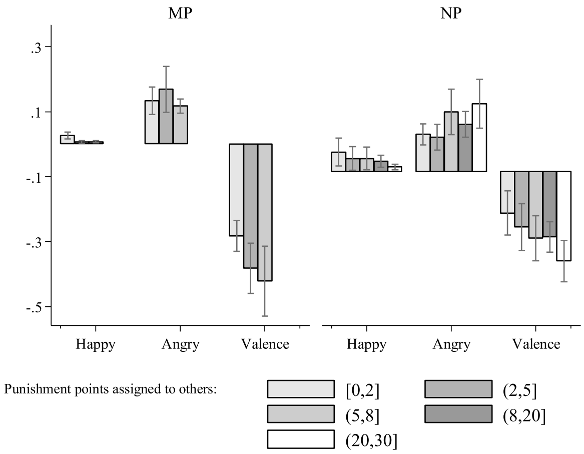

Fig. 3. Happiness, Anger, Valence and the Punishment Assigned to Others. Notes: This figure plots the average emotional states of Happiness, Anger, and Valence during the 10 s before assigning the punishment points and the punishment points assigned to others in five categories. When a participant took less than 10 s to decide, then the average is taken over the shorter available timeframe. Each person could assign up to 10 punishment points to each of the other group members. The error bars indicate the 90 % confidence interval around the mean.

Table 4
Quantity of punishment assigned and emotional state before the punishment decision.

	MP	NP	MP	NP	MP	NP
Angry	1.760	2.322	1.939	−0.484		
	(1.9948)	(4.4749)	(1.8282)	(4.8311)		
Happy	−1.829	−2.394	−3.168	−0.110		
	(4.8448)	(5.3501)	(4.1630)	(6.2118)		
Sad	2.621	11.565*	3.264**	9.044		
	(1.7461)	(6.1744)	(1.5917)	(5.8953)		
Surprised	−0.973	0.899	1.029	5.245		
	(6.0253)	(10.0308)	(4.9611)	(9.6778)		
Scared	3.288	27.385**	4.984	26.022**		
	(8.4517)	(12.1319)	(7.6521)	(12.1670)		
Disgusted	6.960**	9.504	6.165**	17.438*		
	(3.0562)	(10.2676)	(2.7772)	(9.0967)		
Arousal	−2.109	−3.103	−1.368	−4.017		
	(2.4898)	(4.7196)	(2.2486)	(5.0780)		
Period	−0.161**	0.774***	−0.155**	0.759***		
	(0.0761)	(0.1517)	(0.0691)	(0.1582)		
Difference in Contribution			0.397***	0.343***		
			(0.0587)	(0.0588)		
Constant	−1.926	3.664*	−2.359**	4.454**		
	(1.2076)	(2.0966)	(1.1024)	(2.0808)		
Valence					−2.954**	−6.036***
					(1.3155)	(2.2150)
Period					−0.162	0.769***
					(0.0762)	(0.1509)
Constant					−2.429***	4.051***
					(0.7838)	(1.0203)

Notes: Entries are estimated coefficients from random effect Tobit panel regressions for the MP treatment (Tobit is employed because there is a large number of 0 observations in punishment points assigned in the MP treatment); and from random effects panel regression for NP. The dependent variable is the number of punishment points assigned to other people and the explanatory variables are the average emotions ten seconds before assigning punishment points. Standard errors are reported in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Result 3: More negative valence is correlated with a greater assignment of punishment. Assigning punishment increases the valence of the sanctioner. Hypothesis 3 receives partial support.

We now consider the emotional reaction to receiving punishment. The regressions reported in Table 6 provide estimates of the impact of being punished on emotions and establish whether the relationships shown in the figure are significant. The dependent variables are the changes in emotion from the ten seconds preceding the moment in which a subject learns of her punishment for the period, and the ten subsequent seconds. The specifications are of the form:

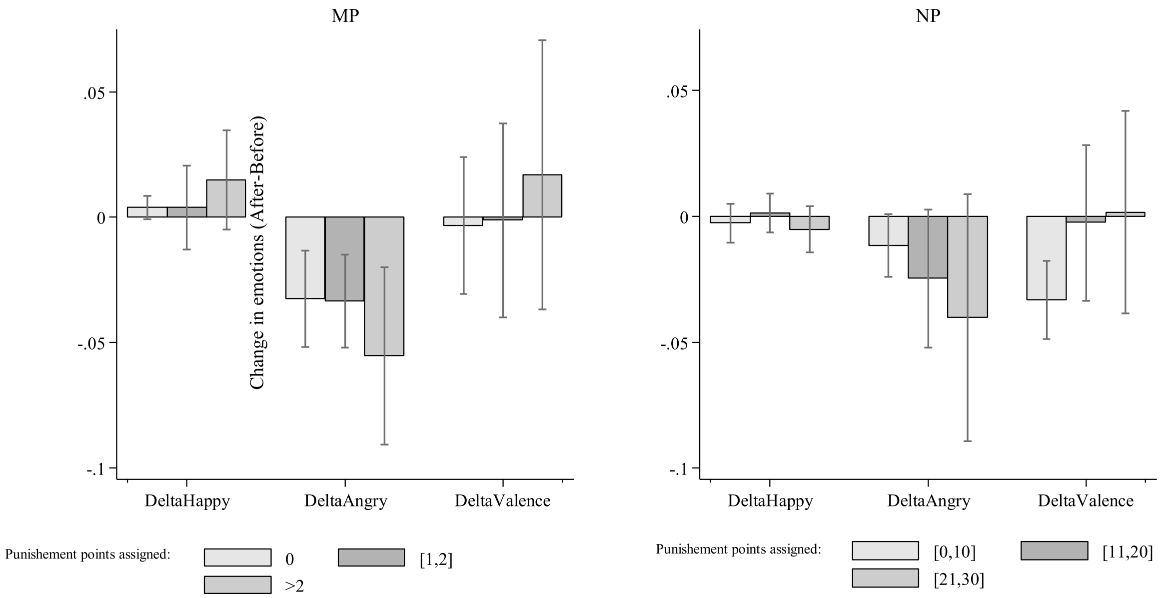

Fig. 4. Emotional Changes After Assigning Punishment to Others. Notes: This figure plots the average change in the emotional states of Happiness and Anger between before and after assigning the punishment points to others. It is calculated as the difference between the average emotion 10 s after assigning the punishment points and 10 s before making the punishment decision. When participants took them less than 10 s to decide on the punishment points, the average is taken over the shorter timeframe available. For the MP treatment, most of the punishment points assigned are 1 or below, and few players (less than 10 %) assigned more than 2 points. This is the reason that the category on the x-axis of the MP figure is different from that of NP.

Table 5
Regressions of the change in emotional state after assigning punishment to others.

	MP	NP	MP	NP	MP	NP
Anger	−0.001 (0.0020)	−0.001 (0.0009)				
Happiness			0.001 (0.0022)	4.00E-05 (0.0004)		
Valence					0.007* (0.0039)	0.002** (0.0011)

Notes: Entries are estimated coefficients of the total punishment points assigned from random effect panel regressions (specification chosen by a Hausman test). The dependent variable is the average change in an emotion from the ten seconds before to the ten seconds after assigning the punishment points to others. Standard errors, clustered at individual level, are reported in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6
Impact of receiving punishment points on changes in emotions.

	MP	NP	MP	NP	MP	NP
Anger	0.003 (0.0033)	0.0007 (0.0006)				
Happiness			0.003 (0.0020)	5.00E-04 (0.0003)		
Valence					0.002 (0.0052)	−0.001** (0.0006)

Notes: Entries are estimated coefficients of the received punishment points from random effect panel regressions (specification chosen with a Hausman test). The dependent variable is the average change in an emotion between ten seconds after and ten seconds before being informed of the punishment points one has received. Standard errors clustered at individual level are reported in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

$$\Delta E_i^t = \beta_0 + \beta_1 \left[\sum_j P_{ji}^t\right]$$

The estimates show that under NP, the more punishment one receives, the larger the decrease in Valence, although the coefficient is very small. There is no significant relationship between the number of points received and the change in Anger and Happiness in either treatment. There is also no relation between the punishment received and the change in Valence under MP. Other regressions (not shown here) demonstrate that the correlation between Valence and sanctions under NP is a consequence of significantly positive correlations between Sadness and Disgust with punishment received. Our findings are summarized as our Result 4.

Result 4: Receiving sanctions generally does not affect Happiness, Anger, or Emotional Valence, with the exception of a small negative effect on Valence in NP. Hypothesis 4 is not supported.

5.4. Emotions and changes in contributions from period to period

In our final analysis, we evaluate how emotions correlate with the change in behavior from one period to the next. We estimate the specification:

$$c_i^t - c_i^{t-1} = \beta_0 + \beta_1 Happy + \beta_2 Angry + \beta_3 Sad + \beta_4 Scared + \beta_5 Disgusted + \beta_6 Surprised + \beta_7 Arousal$$

where the emotions are those in the ten seconds after participants learn their payoff and how much they were punished in period t-1. Additional specifications include $c_i^{t-1} - \frac{\sum_j c_{j\neq i}^{t-1}}{n}$ (Difference in Contribution) and $\frac{\sum_j P_{j\neq i}^{t-1}}{n}$ (Punishment Received) as independent variables. The results are reported in Table 7.

The estimates show that Valence at the end of period t correlates with increasing contributions in period $t+1$ under Baseline and NP, and Happiness does so under Baseline. When there is no monetary punishment involved, the more positive one's emotional state at the end of the last period, the more one contributes. When the difference in contributions between a subject and the average of their group members is included in the estimated equation, this difference correlates with the increase in contributions in $t+1$ and Valence no longer shows a significant relationship. This means that the relationship between Difference in Contribution and the change in cooperation in these two treatments is accounted for by the variation in Valence that results from the Difference in Contribution. This also implies that those who contributed below average tend to have higher Valence.

Under MP, the Difference in Contribution and Punishment Received correlate positively with an increase in contributions. The relationship between these first two variables and the change in cooperation is consistent with previous work. Relatively low (high) contributors increase (decrease) their contributions more, and those who are punished more severely increase their contribution to a greater extent. Thus, monetary punishment is effective in inducing individuals to contribute more. Under NP, receiving points does not change contribution level beyond the effect of the difference in one's own contribution and the mean.

From the Valence analysis in Table 7, we see that people with higher Valence and those who received higher punishment points contribute more in MP. However, it does not show how people with different levels of Valence react to the punishment points received. Table 8 reports the regression analysis that considers the interaction between Valence and punishment points received for both MP and NP. The specifications are of the form:

$$c_i^t - c_i^{t-1} = \beta_0 + \beta_1 Valence_{t-1} + \beta_2 \frac{\sum_j P_{j\neq i}^{t-1}}{n} + \beta_3 Valence_{t-1} * \frac{\sum_j P_{j\neq i}^{t-1}}{n} + \beta_4 \left(c_i^{t-1} - \frac{\sum_j c_{j\neq i}^{t-1}}{n}\right),$$

where $c_i^{t-1} - \frac{\sum_j c_{j\neq i}^{t-1}}{n}$ is the difference in contribution, and $\frac{\sum_j P_{j\neq i}^{t-1}}{n}$ is the amount of punishment points received in the last period. The estimates show that in MP, the interaction term between Valence and the number of punishment points received at the end of period *t-1* correlates with a decrease in contributions in period *t*. This means that the effectiveness of monetary punishment in amending behavior in the following period depends upon the recipient's emotional state. That is, the effectiveness of Punishment decreases for people with higher Valence. The estimates reported in Table 8 show that if the punishment points received are sufficiently high, the overall effect may actually be negative and thus the sanctions may backfire.

6. Discussion and conclusion

In this study, we have measured the emotional correlates of the various decisions taken and types of information received in an interaction consisting of an opportunity to cooperate and then to punish others based on their level of cooperation. The behavioral data is consistent with the previous studies of Isaac and Walker (1988), Andreoni (1988), Fehr and Gaechter (2000), Masclet et al. (2003) and Noussair and Tucker (2005) and many others. In a treatment in which punishment is not possible, cooperation declines over time. When monetary punishment is possible, it is applied predominantly against relatively low contributors and serves to increase cooperation. The use of monetary punishment is relatively stable over time as a social norm of contribution at an approximate level is established. Under non-monetary punishment, sanctions are also assigned to low contributors, and also increase cooperation, though its effects begin to erode with time. More and more non-monetary punishment is applied over time, perhaps in an attempt to sustain eroding cooperation or to express greater disapproval as the group's cooperation declines.

Table 7
Impact of emotions, difference in contribution and punishment received on the change in contribution from period *t*-1 to *t*.

	Baseline	MP	NP	Baseline	MP	NP	Baseline	MP	NP	Baseline	MP	NP
Angry	−2.896	−1.436	−0.562	−0.546	−1.478	0.545						
	(2.2625)	(0.8859)	(1.1397)	(5.1042)	(2.0595)	(3.1132)						
Happy	3.643***	−0.575	6.043	9.871*	1.215	−0.181						
	(1.2125)	(3.8558)	(4.1733)	(5.4261)	(4.1249)	(6.2618)						
Sad	−2.761	0.358	−3.338	−4.904	1.734	−5.322						
	(1.7143)	(1.0909)	(2.3503)	(3.7867)	(1.9849)	(3.7449)						
Surprised	−14.359	4.692	−6.144	−8.074	0.535	−3.694						
	(11.6139)	(6.9395)	(5.9609)	(15.3280)	(5.6089)	(11.4270)						
Scared	10.592	9.810	−4.167	4.995	17.487	−4.427						
	(15.3218)	(6.9977)	(13.5747)	(24.8444)	(12.1038)	(12.3175)						
Disgusted	−1.289	0.590	8.268	−1.270	2.604	−3.858						
	(1.7423)	(1.3809)	(5.1120)	(4.1031)	(3.6861)	(8.1356)						
Arousal	−0.045	4.382*	−1.822	−5.931	−0.515	1.516						
	(3.1104)	(2.5681)	(2.5210)	(4.6595)	(3.0828)	(3.7607)						
Difference in Contribution				−0.671***	−0.705***	−0.748***						
				(0.0783)	(0.1039)	(0.0881)						
Punishment Received					1.841***	0.282						
					(0.6124)	(0.2269)						
Constant	0.233	−1.106	0.430	1.699	−0.621	−0.779						
	(0.6985)	(0.7821)	(1.0339)	(1.7142)	(0.9387)	(1.6210)						
Valence							3.549***	0.809	1.933**	4.837	2.172*	0.594
							(0.8612)	(0.7915)	(0.9657)	(2.9099)	(1.2089)	(1.7442)
Difference in Contribution										−0.668***	−0.448***	−0.633***
										(0.0774)	(0.1102)	(0.0916)
Punishment Received											2.090***	0.019
											(0.5909)	(0.1838)
Constant							0.170	0.565**	0.022	0.462	0.422	−0.378
							(0.2663)	(0.2265)	(0.2784)	(0.7016)	(0.4158)	(0.9114)

Notes: Entries are estimated coefficients from random effect panel regressions. The dependent variable is the change in contribution from one period to the next, $(Cont_t - Cont_{t-1})$, and the explanatory variables are the average emotions during the ten seconds after players learn their payoff and the amount of punishment points received in period *t-1*. Standard errors are reported in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8
The interaction effects of valence and punishment received on the change in contribution between one period and the next.

	MP	NP
Valence	4.03***	0.46
	(1.3782)	(2.5250)
Punishment Received	−0.03	0.03
	(1.0243)	(0.2036)
Valence # Punishment Received	−7.02***	0.06
	(2.5031)	(0.6399)
Difference in Contribution	−0.46***	−0.63***
	(0.0950)	(0.0949)
Constant	0.90	−0.41
	(0.4100)	(0.9968)

Notes: Entries are estimated coefficients from random effect panel regressions. The dependent variable is the change in contribution ($Cont_t - Cont_{t-1}$) and the explanatory variables are the average emotional valence ten seconds after players learn their payoff, the amount of punishment points received in period *t*-1, and the difference in contribution between one's own contribution and other group members' average contributions. Standard errors are reported in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We observe a positive correlation between Happiness just before the contribution decision and the amount contributed in the two punishment treatments. This pattern is similar to that observed by Drouvelis and Grosskopf (2016) who documented a causal relationship between Happiness and cooperation and Joffily et al. (2014) who reported a correlation between self-reported Valence and cooperation. We show here that a similar correlational relationship exists using a physiological rather than a self-reported measure of emotional state and applies to monetary as well as non-monetary symbolic punishment. This correlation between Happiness and contribution does not appear in the Baseline treatment, unlike the results observed by Joffily et al. (2014). One possibility for this lack of a relationship is that when no punishment is possible, attitudes are ambiguous: some individuals are positively disposed to cooperate and increase group payoffs and this behavior accompanies a more positive emotional state, while others are more disposed to free ride and earn greater pecuniary payoffs. This heterogeneity would add up to a lack of overall relationship between emotional states and cooperation. In the MP and NP treatments, those with a disposition to cooperate would feel happier when they do so. However, those who are disposed to free ride would also feel less positive contributing a low amount, as they anticipate the possibility of receiving punishment. The correlation may also suggest that people are happier when they contribute more because they expect less punishment later on. This results in a positive correlation between Happiness and contributions.

Under the Baseline treatment, learning that one has contributed more than others triggers an increase in Anger and a decrease in Valence, which become larger in magnitude as the positive difference between one's own and others' contributions increases. These correlations would be anticipated by previous research (see e.g. Joffily et al., 2014) and are not surprising. However, these relationships do not appear in the MP or NP treatment, and this may be because the upcoming opportunity to punish free riders serves to reduce the anger at being taken advantage of. The absence of an opportunity to punish under the Baseline treatment may be the driving force of the anger that is observed there. While under MP, there is a positive relationship between others' average contribution and one's own Valence, we find no evidence that *mutual* cooperation triggers a more positive subsequent emotional state.

The decision to assign punishment is also linked to emotional states. Under both MP and NP, greater application of punishment is correlated with more negative valence. This result is in agreement with the prior literature (Sanfey et al., 2003; Hopfensitz and Reuben, 2009). However, we find that this relationship is not a consequence of a significant correlation between punishment and either Anger or Happiness. Rather, under both MP and NP, greater *Disgust* is correlated with greater punishment. The results suggest that the emotion of disgust may be a powerful driver of sanctions and it may be the case that in some previous studies negative emotions interpreted as anger may in fact have been registering disgust.[10] Punishment is directed at low contributors, even when we control for emotional states, indicating that not all punishment has emotional correlates. This is perhaps not surprising in view of the fact that punishing others instrumentally can be quite effective in increasing one's own future payoff through its effect on the sanctioned party's contribution. We also find that the more points one assigns, the more the Valence of one's emotional state increases, and the effect is significant in NP. This is consistent with the results reported by de Quervain et al. (2004). We also additionally observe that expressing greater disapproval of other individuals under MP is also associated with an improvement in one's emotional state.

Under MP, receiving punishment is uncorrelated with both changes in Anger and Happiness. This pattern may reflect the reactions to the difference in the number of punishment points received relative to one's prior beliefs about the quantity that they were about to get. Those who receive more than they expected experience an increase in anger and a decrease in happiness, while those who receive fewer points than they anticipated exhibit the opposite pattern, and these effects may offset. Under NP, receiving more punishment points results in a greater decrease in Valence as the effect of being punished more severely than expected has a stronger adverse impact than the positive effect of receiving less punishment than anticipated. In this treatment, the expression of approval or

[10] Sanfey et al. (2003) observe that anterior insula activation is associated with rejecting offers in ultimatum games. Anterior insula activation occurs when an individual is disgusted, as well as when she is angry.

disapproval of others is more salient than in MP, and it appears that disapproval causes one's emotional state to become more negative.

We confirm that the change in an individual's contribution behavior from one period to the next can be predicted by two variables. The first is the amount she has contributed relative to the average, with those who have contributed more than the average lowering their contribution more. The second is the number of punishment points received, with those receiving more points increasing their contribution more, although this is significant only under MP. These patterns were documented by Fehr and Gaechter (2000) and Masclet et al. (2003), among others, and we also observe them here.

If we assume that individuals seek to be in positive emotional states and to avoid negative states, emotional correlates shed light on the effectiveness of punishment instruments. In the absence of punishment, the strongest association between emotions and behavior is that anger and having contributed more than others are correlated. To avoid this anger, a player would seek to contribute less than or equal to the group average, a process that would tend to erode cooperation over time. Indeed, a decline in cooperation over time is the principal pattern observed in a Public Good game when there is no punishment available. When there is no punishment available, there is no consistent relationship between emotional state and cooperation level. We believe that the effect of emotional state on contributions is heterogeneous, with different types of individuals responding in different ways to specific emotions. Parsing out any such relationships must await larger samples and an accumulation of evidence from future research.

On the other hand, when punishment is available, emotional correlates are stronger. Positive emotions accompany cooperation and thus may serve to reinforce cooperative behavior. Free riding is associated with relatively negative emotions. We believe that this is because non-cooperation comes with a risk of being punished, and the anticipation of these emotions may serve to deter free riding. The increase in cooperation that occurs under MP in response to punishment does not appear to rely on its effect on the emotions of either the sanctioner or the recipient of the sanctions. However, the punishment points do have different effects on people with different emotional states. More specifically, punishment points work less effectively for people who are in a more positive emotional state. While individuals in a more negative state are more likely to punish, monetary punishment seems to increase payoffs mainly by changing the relative pecuniary payoffs between cooperation and non-cooperation, and the sanctioner choosing to pay a monetary cost to change the terms of this tradeoff. In addition, the sanctioner experiences some emotional benefit (improved Valence) from punishing.

Under NP, when expressing disapproval is costless, punishing free-riders has the effect of improving the sanctioner's emotional state. The boost in the emotional state of sanctioners encourages the continued use of punishment. Under NP, receiving stronger expressions of disapproval has a more negative impact on emotional state. The impact of receiving non-monetary sanctions on behavior in future periods is indirect. The punishment itself does not influence the subsequent decision to cooperate, nor do emotional states. However, the punishment does serve as a signal of disapproval, which is absent in the baseline treatment. This suggests that emotions have an impact on cooperation in NP by encouraging higher contributions so that the unpleasant emotions associated with the receipt of sanctions can be avoided. It appears that NP, to the extent that it is effective, relies on emotional responses more than a monetary sanctioning scheme. An individual participating in either MP or NP who seeks to maximize the positivity of her emotional state would be a high cooperator and also punish others.

Funding

This work was supported by the University of Waikato Management School.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2023.11.003.

References

Andreoni, J., 1988. Why free ride?: strategies and learning in public goods experiments. J. Public Econ 37 (3), 291–304.
Anderson, C.M., Putterman, L., 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. Games Econ. Behav 54 (1), 1–24.
Andreoni, J., 1995. Cooperation in Public-Goods Experiments: Kindness or Confusion? Am. Econ. Rev. 85 (4), 891–904.
Bartke, S., Bosworth, S.J., Snower, D.J., Chierchia, G., 2019. Motives and comprehension in a public goods game with induced emotions. Theory Decis 86 (2), 205–238.

Battigalli, P., Dufwenberg, M., 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit 60 (3), 833–882.
Bijlstra, G., Dotsch, R., 2011. FaceReader 4 Emotion Classification Performance On Images from the Radboud Faces Database. Unpublished manuscript. , Radboud University Nijmegen, Nijmegen, the Netherlands. Department of Social and Cultural Kugler et al. 323 Psychology.
Bosman, R., van Winden, F., 2002. Emotional hazard in a power-to-take experiment. Econ. J. 112 (476), 147–169.
Breaban, A., Deck, C., Johnson, E., 2022. Emotional differences between isomorphic auctions. J. Behav. Finance 23 (4), 427–437.
Breaban, A., Noussair, C.N., 2018. Emotional state and market behavior. Rev. Financ 22 (1), 279–309.
Carpenter, J.P., 2007. The demand for punishment. J. Econ. Behav. Organ 62 (4), 522–542.
Casari, M., Luini, L., 2009. Cooperation under alternative punishment institutions: an experiment. J. Econ. Behav. Organ 71 (2), 273–282.
den Uyl, M.J., van Kuilenburg, H., 2005. The facereader: online facial expression recognition. In: Proceedings of Measuring Behavior, pp. 589–590.
Denant-Boemont, L., Masclet, D., Noussair, C.N., 2007. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Econ. Theory 33 (1), 145–167.
Dickinson, D.L., Masclet, D., 2015. Emotion venting and punishment in public good experiments. J. Public Econ 122, 55–67.
Drouvelis, M., Grosskopf, B., 2016. The effects of induced emotions on pro-social behaviour. J. Public. Econ 134, 1–8.
Ekman, P., Revealed, E., 2007. Recognizing Faces and Feelings to Improv Communication and Emotional Life. Emotions revealed. New York, NY: Henry Holt.
Elster, J., 1998. Emotions and Economic Theory. J. Econ. Lit 36 (1), 47–74.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Amer. Econ. Rev. 90 (4), 980–994.
Fehr, E., Gächter, S., 2002. Altruistic Punishment in Humans. Nature 415 (10), 137–140.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ 114 (3), 817–868.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett 71 (3), 397–404.
Fiala, F., Noussair, C., 2017. Charitable giving, emotions, and the default effect. Econ. Inq. 55 (4), 1792–1812.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. Amer. Econ. Rev. 100 (1), 541–556.
Galeotti, F., 2015. Do negative emotions explain punishment in power-to-take game experiments? J. Econ. Psychol 49, 1–14.
Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. Cognition. and. Emotion 9 (1), 87–108.
Gürdal, M.Y., Gürerk, Ö., Yahşi, M., 2021. Culture and prevalence of sanctioning institutions. J. Behav. Experimental. Econ. 92, 101692.
Herrmann, B., Thoni, C., Gachter, S., 2008. Antisocial punishment across societies. Science 319 (5868), 1362–1367.
de Hooge, I.E., Zeelenberg, M., Breugelmans, S.M., 2007. Moral sentiments and cooperation: differential influences of shame and guilt. Cogn. Emotion 21 (5), 1025–1042.
Hopfensitz, A., Reuben, E., 2009. The importance of emotions for the effectiveness of social punishment. Econ. J. 119 (540), 1534–1559.
Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. Amer. Econ. Rev. 92 (4), 1062–1069.
Höfling, T.T.A., Gerdes, A.B.M., Föhl, U., Alpers, G.W., 2020. Read my face: automatic facial coding versus psychophysiological indicators of emotional valence and arousal. Front. Psychol. 11, 1388.
Isaac, R.M., Walker, J.M., 1988. Group size effects in public goods provision: the voluntary contributions mechanism. Q. J. Econ 103 (1), 179–199.
Joffily, M., Masclet, D., Noussair, C.N., Villeval, M.C., 2014. Emotions, sanctions, and cooperation. South Econ. J 80 (4), 1002–1027.
Kuderna-Iulian, B., van Kuilenburg, H., Eligio Xolocotzin, U., Den Uyl, M., Cremene, M., Hoszu, A., Octavian, C., 2009. Evaluation of a System for Real-Time Valence Assessment of Spontaneous Facial Expressions. Distributed Environments Adaptability, Semantics and Security Issues International Romanian-French Workshop. Cluj-Napoca.
Kugler, T., Motro, D.J., Noussair, C.N., Ye, B, 2020. On trust and disgust: evidence from face reading and virtual reality. Soc. Psychol. Personal. Sci 11, 317–325.
Ledyard, J.O., 1995. Public Goods: a Survey of Experimental Research. In: Kagel, John H., Roth, Alvin E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ, pp. 111–194, 1995.
Lerner, J., Ye, L., Valdesolo, P., Kassam, K., 2015. Emotion and Decision Making. Annu. Rev. Psychol 66, 799–823.
Lewinski, P., de Uyl, T., Butler, C., 2014. Automated facial coding: validation of basic emotions and FACS AUs in facereader. J. Neurosci., Psychol. Econ. 7, 227–236.
Loewenstein, G., 2000. Emotions in economic theory and economic behavior. Am. Econ. Rev. 90 (2), 426–432.
Loewenstein, G.F., Weber, E.U., Hsee, C.K., Welch, N., 2001. Risk as feelings. Psychol. Bull 127 (2), 267.
Mahmood, M., Gore, C., Kagel, J., 2022. Monetary and Non-Monetary Punishment in Public Goods Games: Comparing Teams With Individuals. Working paper Ohio State University.
Masclet, D., Noussair, C., Tucker, S., Villeval, M.C., 2003. Monetary and nonmonetary punishment in the voluntary contributions mechanism. Amer. Econ. Rev. 93 (1), 366–380.
Matsumoto, D., Willingham, B., 2009. Spontaneous facial expressions of emotion of blind individuals. J. Pers. Soc. Psychol 96, 1–10.
Nelissen, R.M., Zeelenberg, M., 2009. Moral emotions as determinants of third-party punishment: anger, guilt, and the functions of altruistic sanctions. Judgm. Decis. Mak 4 (7), 543–553.
Nguyen, Y., Noussair, C.N., 2022. Incidental emotions and cooperation in a public goods game. Front. Psychol 13.
Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? J. Public. Econ 92 (1–2), 91–112.
Noussair, C., Tucker, S., 2005. Combining monetary and social sanctions to promote cooperation. Econ. Inq 43 (3), 649–660.
Ostrom, E., 1990. Governing the commons: The evolution of Institutions For Collective Action. Cambridge university press.
de Quervain, D.J.F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E., 2004. The neural basis of altruistic punishment. Science 305 (5688), 1254–1258.
Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D., 2002. A neural basis for social cooperation. Neuron 35 (2), 395–405.
Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol 39, 1161–1178.
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game. Science 300 (5626), 1755–1758.
Terzis, V., Moridis, C., Economides, A., 2010. Measuring instant emotions during a self-assessment test: the use of FaceReader. In: Proceedings of Measuring Behavior, pp. 192–195.
Thaler, R.H., 2000. From homo economicus to homo sapiens. J. Econ. Persp. 14 (1), 133–141.
Van Leeuwen, B., Noussair, C.N., Offerman, T., Suetens, S., Van Veelen, M., Van De Ven, J., 2018. Predictably angry—Facial cues provide a credible signal of destructive behavior. Manage Sci. 64 (7), 3352–3364.
Xiao, E., Houser, D., 2005. Emotion expression in human punishment behavior. Proc.. National Acad. Sci. 102 (20), 7398–7401.
Yamagishi, T., 1986. The provision of a sanctioning system as a public good. J. Pers. Soc. Psychol 51 (1), 110.