Games **2010**, *1*, 18-33; doi:10.3390/g1010018

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

Punishment, Cooperation, and Cheater Detection in "Noisy" Social Exchange

Gary Bornstein and Ori Weisel *

Department of Psychology and Center for the Study of Rationality, The Hebrew University,
Mount Scopus, Jerusalem 91905, Israel

* Author to whom correspondence should be addressed; E-Mail: oriw@mscc.huji.ac.il;
Tel.: +972-2-5883006; Fax: +972-2-5881159

*Received: 9 December 2009; in revised form: 1 March 2010 / Accepted: 12 March 2010 /
Published: 15 March 2010*

Abstract: Explaining human cooperation in large groups of non-kin is a major challenge to both rational choice theory and the theory of evolution. Recent research suggests that group cooperation can be explained by positing that cooperators can punish non-cooperators or cheaters. The experimental evidence comes from public goods games in which group members are fully informed about the behavior of all others and cheating occurs in full view. We demonstrate that under more realistic information conditions, where cheating is less obvious, punishment is much less effective in enforcing cooperation. Evidently, the explanatory power of punishment is constrained by the visibility of cheating.

Keywords: public-goods game; punishment; cooperation; reciprocity; experimental games

1. Introduction

The unpredictability of nature is a primary reason why people form groups where they share resources and practice generalized exchange. Foraging societies share meat to reduce the risk inherent in big-game hunting [1–3]. Since even the best hunters face a high risk of failure and provision is on a day-to-day basis, "were each hunter to produce only for his own domestic needs, everyone would eventually perish from hunger" [4]. People in modern societies use similar means for coping with risk [5]. Waiters, whose tips can fluctuate considerably from one day to the next, commonly pool their earnings at the end of the day and then split them evenly. In a typical microfinance scheme [6], borrowers with individual risky projects form groups that apply for loans together and are jointly liable if one or more group members default. Many of us belong to social networks (e.g., extended families, professional or religious communities), where we help others who suffer from some misfortune (e.g., illness, natural disaster) and rely on receiving such help when we are in need [7].

The individual's benefit from being part of such social groups or solidarity networks is clear - a reduction in the variability of personal outcomes and increased protection against the possibility of catastrophic loss. There is also a cost, however, as one is obliged to share with less fortunate others. This payoff structure gives rise to free riding or cheating. While all the members of a group or a society benefit if they all cooperate by sharing their resources, each individual member is better off taking a free ride by withholding resources for private use.[1] How can group cooperation be maintained given this incentive structure?

Recent theoretical work suggests that punishment may provide the answer. It shows that cooperation, even in large groups of non-kin, can be sustained if cooperators can punish non-cooperators [9–15]. Consistent with these theoretical results, laboratory experiments have demonstrated that punishment is indeed highly effective in enforcing cooperation in n-person interactions [16–19]. Typically, in such experiments a group of n individuals plays a public goods game for several identical rounds. In each round, each group member receives a monetary endowment and can contribute any part of it to a common pool. The contributions are then multiplied by a factor larger than 1 but smaller than n, and divided equally among all group members, regardless of their contribution. Thus, in a public goods game, as in the social reality that it models, all group members are better off if they all cooperate by contributing their entire endowment to the common pool, but each group member is better off contributing nothing. By and large, cooperation in this experimental setting has been found to be rather high at the beginning of the interaction but to decline as the game progresses [20,21]. Adding a punishment option dramatically changes these dynamics. When individuals are allowed to use some of their money to punish other group members (after being informed about their contribution), there is an immediate increase in cooperation, followed in time by a further increase to almost full cooperation [16–19].

While these experimental results are very compelling, generalizing them to real-life situations of the type exemplified above is hardly straightforward. In the experiments group members receive commonly known endowments, which are typically identical across group members and fixed over time, and, since contributions are also fully observable, identifying free riders is easy. In real life some group members attain more resources than others, and the distribution of these resources varies over time. Some of the variability, both across group members and over time, results from unobservable random moves of nature (*i.e.*, luck), which render the exact size of one's income at any particular time private information [22]. Not knowing precisely how much the other group members have makes the detection of free riders much more difficult. Is punishment still effective in sustaining group cooperation under these more realistic information conditions?

As a starting point for answering this question, we studied the effect of punishment on cooperation in a repeated public goods game under two information conditions. In both conditions individual

[1] Indeed, a common practice among the Anbara, an Aboriginal tribe where food-sharing norms are strongly enforced, is eating during food collection so that the greater part of a person's take is in an advanced state of digestion by the time he or she returns to camp [8].

endowments in each round of the game were drawn randomly (and independently) from the same, commonly known, distribution. In the private information (PRIVATE) condition each player was informed only about her own endowment. In the public information (PUBLIC) condition the players were also informed about the endowments received by each of the other group members. Comparing these two treatments allowed us to investigate how the asymmetry of information affects punishment and cooperation, while controlling for any potential effects of the mere heterogeneity in endowments among group members.

Previous experiments have demonstrated that income heterogeneity among players has a mostly negative effect on contribution levels in public goods games [21], both linear [23,24] and nonlinear [25,26], but some experiments have found positive effects [27], suggesting that "the evidence on the effect of asymmetric endowments on cooperation levels is far from being conclusive" [22]. There is also a small strand of literature investigating the effect of information about others' endowments on contribution in public goods games. Two experiments using a one-shot step-level public goods game [28,29] found no difference in contribution levels between private and public information conditions. Another study, investigating a repeated linear public goods game, found that information about endowments had no impact on relative contributions unless the group had a leader, in which case incomplete information yielded lower contributions [22].

The rest of the paper is organized as follows. The second section details the experimental procedure; the third section describes the results; and the fourth section summarizes and concludes.

2. The experiment

2.1. Method

Participants

144 undergraduate students (79 females and 65 males) at the Hebrew University of Jerusalem participated in the experiment. Economics students were not excluded. The participants were recruited by campus advertisements promising monetary rewards for participation in a decision-making task.

Experimental procedure

Sessions were held with cohorts of 12 participants. Six cohorts took part in the PRIVATE condition and 6 in the PUBLIC control condition. The experiment was computer-controlled. Upon arrival each participant was seated in a separate cubicle facing a personal computer. The participants were given detailed written and oral instructions explaining the rules and payoffs of the game. The 12 participants in each session were randomly divided into three 4-person groups. Participants were told that they would remain in the same group throughout the experiment, but were not told who the other members of their group were. They were assured that their decisions would be strictly confidential and that at the end of the experiment they would receive their payment one at a time with no opportunity to meet the other participants.[2] Following the last round participants were debriefed about the rationale and purpose of

[2] Although negative payoffs were possible, we assumed that they would be a very unlikely occurrence, given that it was not possible to lose money in the first stage of the game (See subsequent paragraphs). Had any of the participants asked, we

the experiment. Their profits were cashed in at a rate of 1 New Israeli Shekel (NIS) per 4 Money Units (MUs) (1 NIS was equal to about €0.20 at the time the experiment took place). The sessions lasted from one to one and a half hours, and the average earning was around €14. Following payment, participants were dismissed individually.

The game was played by groups of four members in two stages of 18 rounds each. The first stage was played without punishment, while the second included a punishment option. The group composition remained constant throughout the experiment ("partners" design), and the participants were informed about the repeated nature of the task but not about the exact number of rounds, nor that the experiment included a second stage. At the beginning of each round, the four players were allocated their endowments, which were sampled, independently for each player and across rounds, from a flat distribution between 1 and 9 MUs. The distribution and the nature of the sampling process were common knowledge. Each player then indicated the number of MUs, between zero and the total number of MUs allocated to her on that round, that she was wants to contribute to the common pool. The total number of MUs contributed to the pool was multiplied by two and divided equally among the four group members regardless of their individual contributions. Formally, the payoff for participant i in round k without punishment (k ∈{1,2,...,18}) is

$$\Pi_i^k = e_i^k - c_i^k + \frac{1}{2}\sum_{j=1}^{4} c_j^k \tag{1}$$

where e_i^k ∈{1,2,...,9} denotes participant i's endowment in round k and c_i^k ∈{0,1,...,e_i^k} is her contribution in the round.

Following the completion of a round the participants were informed about the number of MUs contributed by each of the other group members (identified by fixed number codes). In the PUBLIC condition the players were also informed about the endowments received by each group member in this round, while in the PRIVATE condition this information was not disclosed. The punishment option was introduced in the second stage of the experiment. Following the feedback, each group member could use up to 5 MUs from what they had earned so far to punish one other group member. For each MU used the punished person would lose 3 MUs [17]. The payoff for participant i in round k with punishment (k ∈{19,20,...,36}) is

$$\Pi_i^k = e_i^k - c_i^k + \frac{1}{2}\sum_{j=1}^{4} c_j^k - p_{i\rightarrow}^k - 3p_{i\leftarrow}^k \tag{2}$$

where $p_{i\rightarrow}^k$ denotes the number of MUs participant i allocated to punishing another participant in round k and $p_{i\leftarrow}^k$ denotes the number of MUs other participants used to punish i in round k. After all players had made their punishment decisions, they received additional feedback about the number of MUs, if any, they had used for punishing; the number of MUs, if any, they had lost due to being punished by others; and their final earnings in that round. The identity of the punishing group member was not disclosed.

would have assured them privately that they would not be asked to pay out any money. However, none of the participants asked, and nobody ended up with a negative payoff.

3. Results

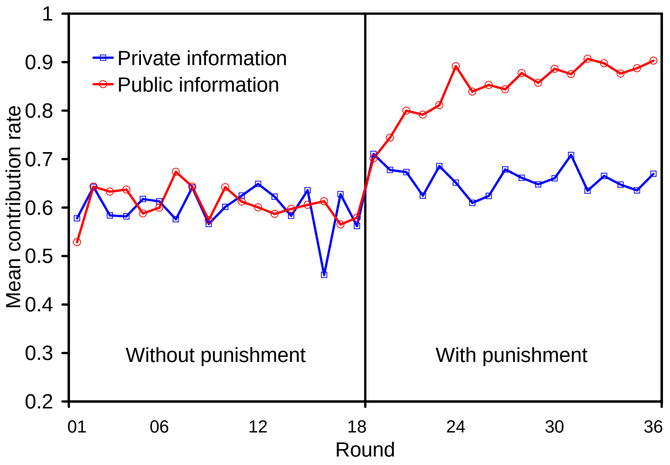
Contribution rates

Our first analysis refers to contribution rates - the proportion of the endowments that group members chose to contribute out of the total endowment they received in a particular round. For a given group, the contribution rate in round k (CR^k) is defined as

$$CR^k = \frac{\sum_{i=1}^{4} c_i^k}{\sum_{i=1}^{4} e_i^k}. \tag{3}$$

In the first (18-round) stage of the game played without punishment there was virtually no difference in contribution rates between the PUBLIC and PRIVATE conditions. The average contribution rate was 0.61 (SD=0.25, based on 18 time-averaged group contribution rates) and 0.60 (SD=0.14) in the two conditions, respectively. The introduction of the punishment option in the game's second (18-round) stage increased the average contribution rate in the PUBLIC condition to 0.85 (SD=0.16). All 18 groups in the PUBLIC condition contributed more in the punishment stage than in the preceding no-punishment stage. In the PRIVATE condition punishment had a more modest effect, increasing the average contribution rate to only 0.66 (SD=0.17). Thirteen of the 18 groups in this condition contributed more in the punishment stage than in the no-punishment stage. The mean contribution rates per round are presented in Figure 1.

Figure 1. Mean contribution rates for each round. In the first stage of the game (rounds 1-18), played without punishment, cooperation rates were nearly identical in both conditions. In the second stage (rounds 19-36), when punishment was made possible, the cooperation rate in the PUBLIC condition was higher than in the PRIVATE condition.



We used a Wilcoxon signed rank (WSR) test with 18 matched observations in each condition to compare contribution rates with and without punishment in the two information conditions. Each matched observation consisted of a group's average contribution rate with punishment and its average contribution rate without punishment. The analysis yielded S=85.5, p<.0001, for the PUBLIC condition and S=48.5, p=.0171, for the PRIVATE condition (both one-tailed)[3]. A Wilcoxon rank-sum (WRS) test was used to examine the difference in contribution rates between the PUBLIC and PRIVATE conditions in the punishment stage, and yielded U=263 (n_1=n_2=18), Z=3.18, p=.0007.[4]

Punishment behavior

Punishment was used somewhat more frequently in the PUBLIC condition than in the PRIVATE condition. In the PUBLIC condition punishment was used at least once in 17 of the 18 groups (94%), as compared with only 14 of the 18 groups (78%) in the PRIVATE condition. To test this difference, each group was a given a score of 1 if punishment was used by any of its members at least once, and a score of 0 otherwise. A WRS test on these binary scores yielded U=189 (n_1=n_2=18), Z=1.40, p=.081. Forty-three participants (60%) punished others at least once in the PUBLIC condition, as compared with 36 participants (50%) in the PRIVATE condition. As in the previous test, a score of 1 was assigned to participants who chose to punish someone else at least once, and a score of 0 to participants who never punished anyone. A WRS test on these scores yielded U=2844 (n_1=n_2=72), Z=1.17, p=.122. Finally, more MUs were used for punishment in the PUBLIC condition (5.5 per player) than in the PRIVATE condition (3.36).[5] A WRS test was used to examine this difference, resulting in U=206.5 (n_1=n_2=18), Z=1.40, p=.082, for a group level analysis and U=3017.5 (n_1=n_2=72), Z=1.78, p=.037, for an individual level analysis.

Who was more likely to punish or to be punished?

In each 4-person group we identified the most and least cooperative members, based on their relative contributions in the first (no punishment) stage of the game, and compared the extent to which these players punished others or were punished themselves during the game's second stage.[6] In the PUBLIC condition, as can be seen in Figure 2a, the most cooperative group members were more likely than the least cooperative ones to punish others at least once (WRS test: U=108, (n_1=n_2=18), Z=-1.97, p=.025), and used significantly more MUs for this purpose (U=94.5, (n_1=n_2=18), Z=-2.22, p=.013). The most cooperative members were also significantly less likely than the least cooperative ones to be punished at least once (U=207, (n_1=n_2=18), Z=1.76, p=.039) and were punished by significantly fewer MUs (U=243, (n_1=n_2=18), Z=2.59, p=.005).

This pattern of punishment significantly affected the variability of the within-group payoffs. We computed, for each group, the standard deviation of its members' payoffs in the first (no punishment) and second (punishment) stages of the game. The resulting standard deviations were smaller on average

[3] Unless otherwise stated, reported p-values are one-tailed.

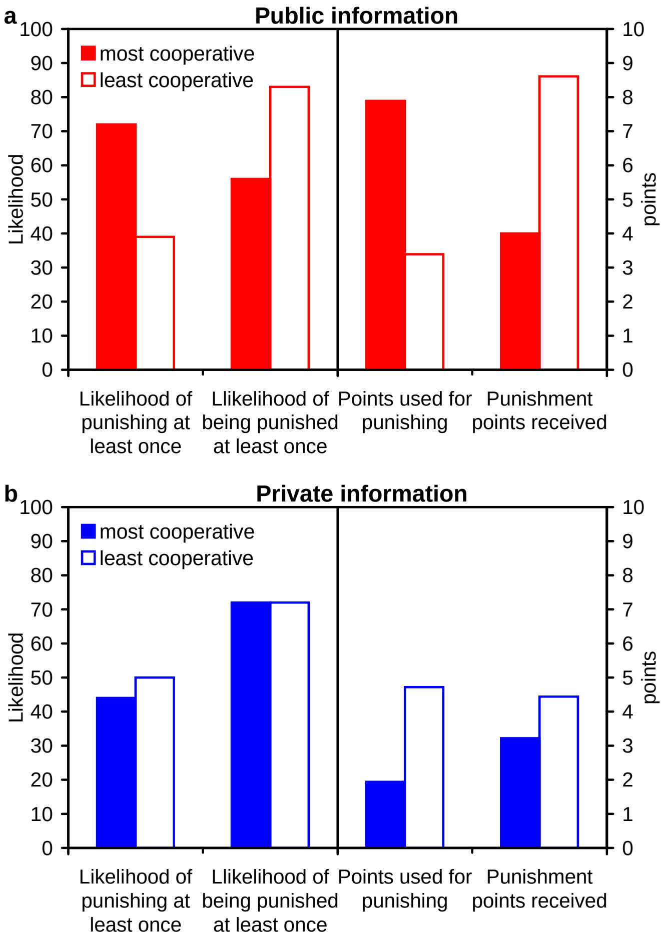
[4] The Z score, here and in all WRS tests in the paper, includes a continuity correction of 0.5.

[5] SDs, treating individuals as the unit of analysis, are 15.55 for the PUBLIC condition and 11.31 for the PRIVATE condition. Treating groups as the unit of analysis yields 5.6 and 3.47, respectively.

[6] The most cooperative members contributed, on average, 75% of their endowments in the first stage of the game in the PUBLIC condition and 74% in the PRIVATE condition. The contribution averages for the least cooperative members were 44% and 45%, respectively.

in the punishment stage (8.81 MUs) than in the no-punishment stage (13.9 MUs) (WSR Test: S=-46.5, p=.021. Each matched observation consisted of the two SDs computed for each group).[7]

Figure 2. Punishment behavior of the most and the least cooperative group members. a - In the PUBLIC condition the most cooperative group members were more likely to engage in punishment and less likely to have punishment directed at them (left side of figure). They also used more MUs for punishment and had less punishment directed at them than the least cooperative group members (right side). **b** - This was not the case in the PRIVATE condition.



These relations between contribution and punishment, which are similar to those found in previous experiments [17,31], were not observed in the PRIVATE condition (Figure 2b). The most and the least cooperative group members did not differ significantly in their willingness to punish others or the likelihood that they would be punished themselves. They also did not differ significantly in the number of MUs they used for punishing others, or the number of MUS used to punish them.[8] Consequently, punishment in the PRIVATE condition did not have a significant effect on the within-group payoff variation

[7] A similar effect of punishment on the re-distribution of wealth was reported by Visser [30].

[8] No statistical tests are provided to establish the differences in the likelihood of punishing others, the likelihood of being punished, or the number of MUs used for punishing others in the PRIVATE condition because these differences either did not exist or were in the opposite direction from the (significant) differences found in the PUBLIC condition. The difference in the

(average standard deviations of group members' payoffs were 12.36 in the punishment stage and 13.34 in the no-punishment stages; WSR test: S=-12.5, p=.305).

Punishment and deviation from group contribution

The analysis presented in the previous section associated group members' behavior in the first stage of the game, played without punishment, with the overall punishment they inflicted and received in the second stage. To further investigate the dynamics of punishment and cooperation in the second stage, we tested the association between the number of punishment MUs given to group member i in a given round (k) and the (negative or positive) deviation of her contribution from the contributions of the others in her group. This association was tested in terms of both the participants' *relative* (c_i^k/e_i^k) and *absolute* (c_i^k) contributions. We first consider the deviations of i's *relative* contribution from that of the other group members. We used Tobit regressions, taking into account that only observations across groups are independent, with punishment as a censored variable. In the PUBLIC condition the regression coefficients are 12.32 (Z=14.31, p<.0005) on negative deviations and 3.42 (Z=2.30, p=.021) on positive deviations. In the PRIVATE condition the coefficients are 3.45 (Z=1.93, p=.053) on negative deviations and 3.37 (Z=2.53, p=.011) on positive deviations. The interpretation of these coefficients is straightforward; an increase of 0.1 in the negative deviation of i's contribution rate from the contribution rate of i's group is associated with an increase of 1.232 in the number of MUs allocated to punish i in the PUBLIC condition, but only 0.345 MUs in the PRIVATE condition. The effect of positive deviations is similar in both conditions. An increase of 0.1 in the positive deviation resulted in an increase of 0.342 MUs in punishment in the PUBLIC condition and 0.337 in the PRIVATE condition.

We conducted the same analysis on the association between the deviations of *absolute* contributions from the average contribution of other group members. The analysis yielded nonsignificant values for both negative (Z=.57, p=.571) and positive (Z=.-.14, p=.892) deviations in the PUBLIC condition. In the PRIVATE condition, however, participants were punished for negative deviations in absolute contributions (coefficient: 1.173, Z=4.55, p<.0005). Positive deviations did not have a significant effect on punishment in this condition (coefficient: 0.278, Z=.97, p=.332).

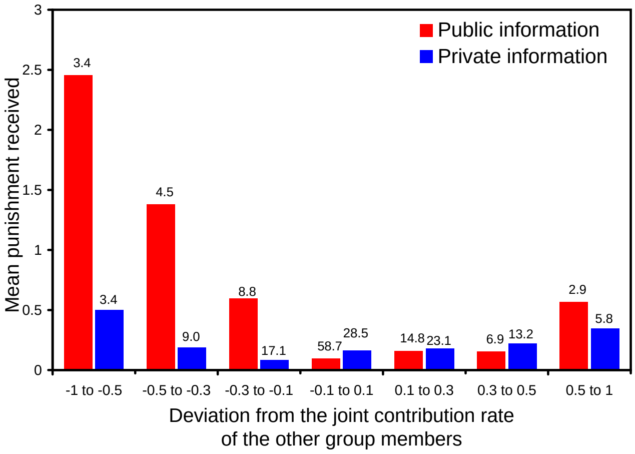
Effect of punishment on lagged contributions

The contribution rate analysis above demonstrated that punishment had a greater positive effect on cooperation in the PUBLIC condition than in the PRIVATE condition. The following analysis further explores the effect of punishment on contribution in the two conditions, by investigating the direct effect of punishment in a given round on contribution rates in the next round. For this we checked whether punishment i received in round k had an effect on i's contribution rate in round $k+1$. Again, we used Tobit regressions, taking into account that only observations across groups are independent, this time with the difference in the contribution rates in rounds k and $k+1$ as a censored dependent variable.[9] The analysis yielded a significant coefficient of 0.12 (Z=5.09, p<.0005) in the PUBLIC condition, meaning

number of punishment MUs received, while in the predicted direction, was not significant (U=181.5, (n_1=n_2=18), Z=0.61, p=.271).

[9] The dependent variable is $(\frac{c_i^{k+1}}{e_i^{k+1}} - \frac{c_i^k}{e_i^k})$.

Figure 3. Punishment as a function of deviation from the group contribution rate. The number of punishment MUs that were directed at player i as a function of i's deviation from the contribution rate of the other group members. Negative deviations from the others' contribution rate were associated with more punishment in the PUBLIC condition, but not in the PRIVATE condition. The numbers above the bars indicate the relative frequency of the relevant observations.



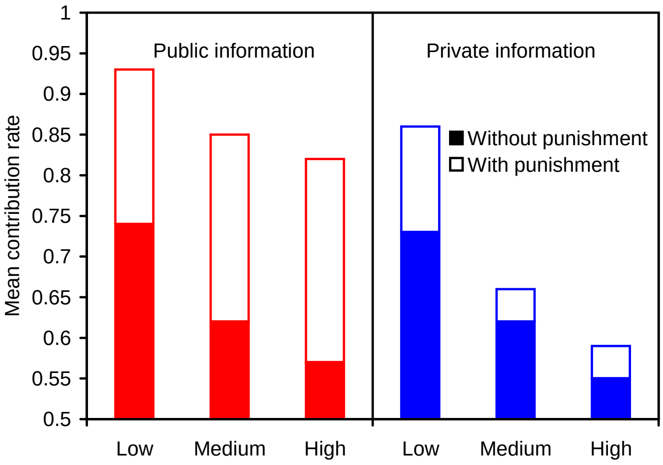
that each MU of punishment for i increased her contribution rate in the next round by 12%. The same coefficient in the PRIVATE condition is not significant (Z=-1.38, p=.166). Considering only whether i was punished or not in the previous round (*i.e.*, ignoring the amount of the punishment) yields a similar result. The coefficient is 0.42 (Z=6.45, $p<$.0005) in the PUBLIC condition, and nonsignificant (Z=-1.29, p=.198) in the PRIVATE condition.

Endowment effect

We also investigated whether the effect of punishment on contribution depends on one's endowment in a particular round. This was done by comparing the average contribution rates with and without punishment separately for low (1-3), medium (4-6), and high (7-9) endowment levels. Mean contribution rates with and without punishment for the three endowment levels are shown in Figure 4. In the PUBLIC condition, punishment significantly increased contribution rates for all endowment levels. A WSR test for 18 matched observations yielded S=68, S=85.5, and S=83.5 for low, medium and high endowments respectively ($p<$.0001 for all three). In the PRIVATE condition, punishment had a similar effect on contribution rates for low endowments (S=80.5, $p<$.0001), but a smaller effect for medium and high endowments (S=38.5, p=.049; S=24.5, p=.150).[10]

[10] The analysis was done at the group level. For each group we computed the average contribution rate of its members for each endowment level, separately for the no-punishment and punishment stages. A matched observation consists of these two averages.

Figure 4. Contribution rates according to low (1-3), medium (4-6), and high (7-9) endowments. In the PUBLIC condition punishment increased contribution rates for all endowment levels. In the PRIVATE condition punishment increased contribution for low endowments, but had little effect on the contribution of players with medium or high endowments.



Collective efficiency

Increasing the contribution rate is not the ultimate goal of a punishment mechanism. Since punishment is costly, both for those who punish others and even more so for those who are punished, it can potentially decrease the group's collective welfare even though individual contributions increase [32]. In the present experiment, group members could have doubled their joint earnings had they all contributed their entire endowment and completely avoided using punishment. Specifically, we define a group's efficiency in round k (EF^k) as the ratio between the sum of the group members' actual payoffs in that round and the sum of their maximal possible joint payoff:[11]

$$EF^k = \frac{\sum\limits_{i=1}^{4} \Pi_i^k}{2\sum\limits_{i=1}^{4} e_i^k}. \tag{4}$$

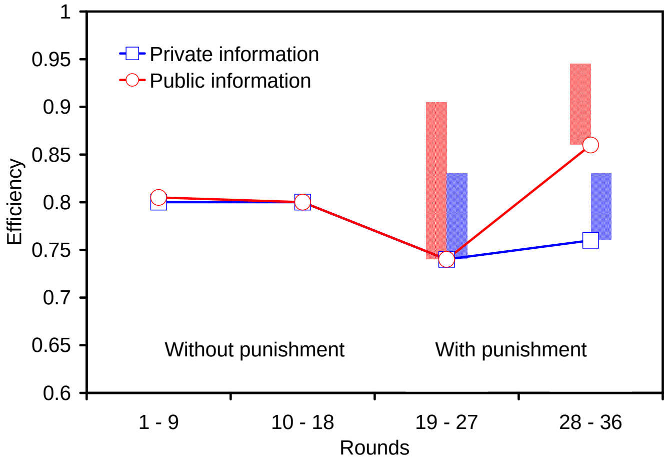
Comparing the average efficiency in the first and second stages of the game reveals that punishment did not increase group efficiency in either the PUBLIC or the PRIVATE condition. Nevertheless, when punishment was allowed, the average efficiency in the PUBLIC condition was significantly higher than that in the PRIVATE condition (0.80 (SD=0.20) and 0.75 (SD=.13), respectively; WRS test: U=221, (n_1=n_2=18), Z=1.85, p=.032).[12] The temporal pattern of efficiency change, presented in Figure 5, might be more informative. The figure depicts the mean efficiency scores in the first and second (9-trial) halves of each of the game's two stages. As can be seen in the figure, punishment initially reduced collective

[11] The maximal possible joint payoff for a group is simply twice the sum of group members' endowments. This happens when all group members contribute their entire endowment, and no MUs are used for punishment.

[12] SDs are based on 18 time-averaged group efficiency scores.

efficiency in both conditions. However, in the PUBLIC condition there was a large and statistically significant efficiency gain in the second half of the punishment stage (0.74 in the first half vs. 0.86 in the second; WSR test with 18 matched observations: S=65.5, p=.0004), whereas in the PRIVATE condition no such gain was observed (0.74 vs. 0.76, S=18.5, p=.442).

Figure 5. Mean efficiency and efficiency loss for each 9-round block. The maximal group efficiency is 1. The solid lines connect the efficiency values and the vertical bars represent the efficiency loss due to punishment. If participants could have managed the same contribution rates without using any punishment, efficiency values would be at the top of the vertical bars.



4. Discussion and implications

When group members are fully informed about the endowments and contributions of all others, providing them with a punishment option increases cooperation levels, as demonstrated in several public goods experiments [17,33,34]. This effect of punishment is explained by the presence of strong reciprocators who enforce the norm of conditional cooperation by cooperating with cooperative others and punishing the non-cooperative ones [12,35–37]. The results of the PUBLIC condition of the present study are clearly in line with these earlier results. Furthermore, the present study adds to this literature by showing that punishment remains effective even when group members' endowments in each period are determined by chance, as long as these endowments are known and cheating occurs in plain sight.

However, in most real-life social exchange situations, where individual resources fluctuate unpredictably, cheating is not so readily detectable. The PRIVATE condition demonstrated that punishment opportunities have little effect on the attainment of mutual cooperation when information about individual endowments is incomplete. Punishment is much less effective in the PRIVATE condition because group members are more reluctant to use this option when information is scant, and, more critically, when punishment is used it is largely misdirected. As a result, punishment fails to discipline selfish group members, or to assure the norm-abiding ones that the selfish players will also cooperate.[13]

[13] The overall use of punishment in our experiment was lower than that observed in other public goods experiments (e.g., [17]). This perhaps can be explained by the fact that the endowments in the present experiment were fairly low (an average

Of course, it remains to be seen whether cheater detection is more reliable and punishment is more effective when resources are drawn from more realistic and more informative distributions, such as normal distributions, where very high or low endowments are unlikely and group members receive intermediate size endowments most of the time. Additionally, cheater detection should improve in longer interactions where group members can gather more information about each other.

There are additional issues involving the generalizability of our experimental results to real-life situations of social exchange that need to be discussed. In the experiment all the group members drew their endowments from the same distribution, and the differences among them were due solely to luck. In real life, some group members are better hunters or better waiters than others, and can, on average, secure greater resources. Keeping one's true ability as private information opens up additional opportunities for free riding; by hiding their true type, high-ability individuals can keep a greater portion of their resources without being detected and punished. The current investigation also ignores the possibility of free riding via a reduction of one's effort level [38]. When monitoring is imperfect and the actual level of individual effort is not readily observable, group members who exert less effort can free-ride on the effort of others [39]. Future research could incorporate these variables into the experimental paradigm, moving it closer to real-life social exchange situations where group members' resources are determined jointly by their ability, effort and luck.

Like most other public goods experiments, our study (and the extensions suggested above) provided group members with quantitative information about each other's behavior. In the PUBLIC condition subjects were informed about each other's contributions and endowments, and could use this information to identify those who failed to contribute their fair share. In the PRIVATE condition, subjects knew the distribution from which endowments were sampled, but could only estimate the others' relative contributions through repeated interactions [22]. Subjects' success in detecting cheaters in this setting depends on their proficiency as "intuitive statisticians" [40]. In many natural settings, however, group members are likely to have other cues that can be used to detect cheaters. Recent research suggests that individuals have some ability to distinguish between cooperative and non-cooperative individuals based on visual cues. Observers were better able to remember the faces of cheaters than the faces of cooperators when asked to memorize the faces of target persons who had played a prisoner's dilemma game earlier [41], and were able to successfully identify targets who had played cooperatively in the PD game, based on pictures that were taken at the decision-making moment [42]. Moreover, individuals have been shown to be able to identify altruistic traits in others [43]. Based on short video clips of the target persons recorded in a setting unrelated to altruistic behavior, participants were able to predict the behavior of these targets in the dictator game significantly better than chance.

The achievement of mutual cooperation in social exchange is one of the most important adaptive tasks humans have faced throughout their evolution. Given the importance of this task, Cosmides and Tooby

of 5 MUs per round as compared with 20 MUs in most other experiments), making punishment a very potent weapon, which subjects might be reluctant to use. Nevertheless, in the PUBLIC condition punishment was highly effective, increasing cooperation to the same level as in other experiments. Clearly the relative severity of punishment and the fact that it was cautiously used did not hinder its effectiveness. In the PRIVATE condition participants still expended on punishment about 60% of what they did in the PUBLIC condition - a substantial proportion - with a small, arguably negligible, effect on cooperation. The severity of punishment, thus, seems an unlikely explanation for its differential effect in the two conditions. This, however, is essentially an empirical matter that could be studied in future experiments.

[44,45] speculated that humans must have acquired a cognitive module designed specifically for processing information relevant for cheater detection. Much of the experimental work on cheater detection has been conducted in the context of a non-interactive logical task.[14] Little has been done to test cheater detection in the more relevant context of interactive n-person public goods games. The experimental research on human cooperation in public goods games, and in particular that on punishment, typically provided group members with complete information about each other, making their cheater detection skills, if they indeed exist, quite irrelevant. Studying the effect of punishment in public goods games with incomplete information can help close the gap between these two lines of research. By systematically manipulating the type and amount of information available to group members in this prototypical model of social exchange, one can test their ability to detect cheaters where it really matters, and by better understanding this cognitive ability and its limitations, determine the boundaries of strong reciprocity and punishment as explanations of human cooperation.

Acknowledgements

This research was supported by a grant from the U.S.-Israel Binational Science Foundation (2003299). We thank Carlos Alós-Ferrer and two anonymous referees for many helpful comments. We also thank Ido Erev, Ernst Fehr, Simon Gächter, Uri Gneezy, Rosemary Nagel and Eyal Winter for their help.

References

1. Cashdan, E.A. Coping with Risk: Reciprocity Among the Basarwa of Northern Botswana. *Man* **1985**, *20*.
2. Bird, R.B.; Bird, D.W.; Smith, E.A.; Kushnick, G.C. Risk and Reciprocity in Meriam Food Sharing. *Evol. Hum. Behav.* **2002**, *23*, 297–321.
3. Hawkesa, K.; O'Connella, J.; Jones, N.B. Hadza Meat Sharing. *Evol. Hum. Behav.* **2001**, *22*, 113–142.
4. Ingold, T. The Significance of Storage in Hunting Societies. *Man* **1983**, *18*, 517–553.
5. Robson, A.J.; Kaplan, H.S. Viewpoint: The Economics of Hunter-Gatherer Societies and the Evolution of Human Characteristics. *Can. J. Economics* **2006**, *39*, 375–398.
6. Abbink, K.; Irlenbusch, B.; Renner, E. Group Size and Social Ties in Microfinance Institutions. *Econ. Inq.* **2007**, *44*, 614–628.
7. Bloch, F.; Genicot, G.; Ray, D. Informal Insurance in Social Networks. *J. Econ. Theory* **2008**, *143*, 36–58.
8. Hiatt, L., Traditional Attitudes to Land Resources. In *Aboriginal Sites: Rites and Resource Development*; Berndt, R.M., Ed. University of Western Australia Press: Perth, 1982.

[14] This research used the Wason selection task which presents subjects with a conditional rule of the form "If P, then Q" and four two-sided cards, each of which has a P or a Q on one side and a *not* P or a *not* Q on the other. The subjects' task is to select the cards that must be turned over to determine whether the rule has been violated. Numerous studies have found that most people fail to make the logically correct selections in this abstract context [46], but performance is greatly improved in a cheater detection context [44]. Whether these findings prove the existence of a special cheater detection module is highly controversial, however [47–53].

9. Boyd, R.; Gintis, H.; Bowles, S.; Richerson, P. The Evolution of Altruistic Punishment. *P. Natl. Acad. Sci. USA* **2003**, *100*, 3531–3535.
10. Fehr, E.; Fischbacher, U. The Nature of Human Altruism. *Nature* **2003**, *425*, 785–791.
11. Gardner, A.; West, S.A. Cooperation and Punishment, Especially in Humans. *Am. Nat.* **2004**, *164*, 753–764.
12. Gintis, H. Strong Reciprocity and Human Sociality. *J. Theor. Biol.* **2000**, *206*, 169–179.
13. Gintis, H. The Hitchhiker's Guide to Altruism: Gene-Culture Coevolution, and the Internalization of Norms. *J. Theor. Biol.* **2003**, *220*, 407–418.
14. Sigmund, K.; Hauert, C.; Nowak, M.A. Reward and Punishment. *P. Natl. Acad. Sci. USA* **2001**, *98*, 10757–10762.
15. Henrich, J.; Boyd, R. Why People Punish Defectors: Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas. *J. Theor. Biol.* **2001**, *208*, 79–89.
16. Boyd, R.; Richerson, P. Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups. *Ethol. Sociobiol.* **1992**, *13*, 171–195.
17. Fehr, E.; Gächter, S. Altruistic Punishment in Humans. *Nature* **2002**, *415*, 137–140.
18. Gurerk, O.; Irlenbusch, B.; Rockenbach, B. The Competitive Advantage of Sanctioning Institutions. *Science* **2006**, *312*, 108–111.
19. Rockenbach, B.; Milinski, M. The Efficient Interaction of Indirect Reciprocity and Costly Punishment. *Nature* **2006**, *444*, 718–723.
20. Dawes, R.M.; Thaler, R.H. Anomalies: Cooperation. *J. Econ. Psychol.* **1988**, *2*, 187–197.
21. Ledyard, J.O. Public Goods: A Survey of Experimental Research. In *Handbook of Experimental Economics*; Kagel, J.; Roth, A., Eds. Princeton University Press: Princeton, NJ, 1995, pp. 111–195.
22. Levati, V.; Sutter, M.; van der Heijden, E. Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information. *J. Conflict Resolut.* **2007**, *51*, 793–818.
23. Cardenas, J.C. Real Wealth and Experimental Cooperation: Experiments in the Field Lab. *Journal of Development Economics* **2003**, *70*, 263–289.
24. Cherry, T.L.; Kroll, S.; Shogren, J.F. The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab. *J. Econ. Behav. Organ.* **2005**, *57*, 357–365.
25. Rapoport, A.; Suleiman, R. Incremental Contribution in Step-Level Public Goods Games with Asymmetric Players. *Organ. Behav. Hum. Dec.* **1993**, *55*, 171–194.
26. Aquino, K.; Steisel, V.; Kay, A. The Effects of Resource Distribution, Voice, and Decision Framing on the Provision of Public Goods. *J. Conflict Resolut.* **1992**, *36*, 665–687.
27. Chan, K.S.; Mestelman, S.; Moir, R.; Muller, R.A. Heterogeneity and the Voluntary Provision of Public Goods. *Exp. Econ.* **1999**, *2*, 5–30.
28. van Dijk, E.; Grodzka, M. The Influence of Endowment Asymmetry and Information Level on the Contribution to a Public Step Good. *J. Econ. Psychol.* **1992**, *35*, 329–342.
29. van Dijk, E.; Wilke, H.; Wilke, M.; Metman, L. What Information Do We Use in Social Dilemmas? Environmental Uncertainty and the Employment of Coordination Rules. *J. Exp. Soc. Psychol.* **1999**, *35*, 109–135.

30. Visser, M. Welfare Implications of Peer Punishment in Unequal Societies. Working Papers in Economics 218, Göteborg University, Department of Economics, 2006.
31. Fehr, E.; Gächter, S. Cooperation and Punishment in Public Goods Experiments. *Am. Econ. Rev.* **2000**, *90*, 980–994.
32. Dreber, A.; Rand, D.G.; Fudenberg, D.; Nowak, M.A. Winners Don't Punish. *Nature* **2008**, *452*, 348–351.
33. Ostrom, E.; Walker, J.; Gardner, R. Covenants With and Without a Sword: Self-Governance is Possible. *Am. Polit. Sci. Rev.* **1992**, *86*, 404–417.
34. Yamagishi, T. The Provision of a Sanctioning System as a Public Good. *J. Pers. Soc. Psychol.* **1986**, *51*, 110–116.
35. Fehr, E.; Fischbacher, U.; Gächter, S. Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms. *Hum. Nature* **2002**, *13*, 1–25.
36. Gintis, H.; Bowles, S.; Boyd, R.; Fehr, E. Explaining Altruistic Behavior in Humans. *Evol. Hum. Behav.* **2003**, *24*, 153–172.
37. Bowles, S.; Gintis, H. The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations. *Theor. Popul. Biol.* **2004**, *65*, 17–28.
38. Muehlbacher, S.; Kirchler, E. Origin of Endowments in Public Good games: The Impact of Effort on Contributions. *J. Neurosci. Psychol. Econ.* **2009**, *2*, 59–67.
39. Barkan, R.; Erev, I.; Zinger, E.; Tzach, M. Tip Policy, Visibility and Quality of Service in Cafes. *Tourism Econ.* **2004**, *10*, 449–462.
40. Peterson, C.R.; Beach, L.R. Man as an Intuitive Statistician. *Psychol. Bull.* **1967**, *68*, 29–46.
41. Yamagishi, T.; Tanida, S.; Mashima, R.; Shimoma, E.; Kanazawa, S. You Can Judge a Book by its Cover: Evidence that Cheaters May Look Different From Cooperators. *Evol. Hum. Behav.* **2003**, *24*, 290–301.
42. Verplaetse, J.; Vanneste, S.; Braeckman, J. You Can Judge a Book by its Cover: the Sequel: A Kernel of Truth in Predictive Cheating Detection. *Evol. Hum. Behav.* **2007**, *28*, 260–271.
43. Fetchenhauer, D.; Groothuis, T.; Pradel, J. Not Only States But Traits – Humans Can Identify Permanent Altruistic Dispositions in 20 s. *Evol. Hum. Behav.* **2010**, *31*, 80–86.
44. Cosmides, L. The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task. *Cognition* **1989**, *31*, 187–276.
45. Tooby, J.; Cosmides, L., The Psychological Foundations of Culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*; Barkow, J.; Tooby, J.; Cosmides, L., Eds. Oxford University Press: New York, 1992, chapter 1, pp. 1–72.
46. Wason, P.C. Reasoning About a Rule. *Q. J. Exp. Psychol.* **1968**, *20*, 273–281.
47. Sperber, D.; Cara, F.; Girotto, V. Relevance Theory Explains the Selection Task. *Cognition* **1995**, *57*, 31–95.
48. Sperber, D.; Girotto, V. Use or Misuse of the Selection Task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition* **2002**, *85*, 277–290.
49. Fodor, J. Why We Are So Good at Catching Cheaters. *Cognition* **2000**, *75*, 29–32.

50. Fiddick, L.; Cosmides, L.; Tooby, J. No Interpretation Without Representation: the Role of Domain-Specific Representations and Inferences in the Wason Selection Task. *Cognition* **2000**, *77*, 1–79.
51. Atran, S. A Cheater-Detection Module? Dubious Interpretations of the Wason Selection Task and Logic. *Evolution Cognition* **2001**, *7*, 187–193.
52. Beaman, C.P. Why Are We Good at Detecting Cheaters? A Reply to Fodor. *Cognition* **2002**, *83*, 215–220.
53. Fiddick, L.; Erlich, N. Giving It All Away: Altruism and Answers to the Wason Selection Task. *Evol. Hum. Behav.* **2010**, *31*, 131–140.