Nosenzo, Daniele and Sefton, Martin (2014) Promoting cooperation: the distribution of reward and punishment power. In: Reward and punishment in social dilemmas. Series in human cooperation . Oxford University Press, Oxford, pp. 87-114. ISBN 9780199300730

Promoting Cooperation:
The Distribution of Reward and Punishment Power

by

Daniele Nosenzo* and Martin Sefton*

07 January 2013

Abstract

Recent work in experimental economics on the effectiveness of rewards and punishments for promoting cooperation mainly examines decentralized incentive systems where all group members can reward and/or punish one another. Many self-organizing groups and societies, however, concentrate the power to reward or punish in the hands of a subset of group members ('central monitors'). We review the literature on the relative merits of punishment and rewards when the distribution of incentive power is diffused across group members, as in most of the extant literature, and compare this with more recent work and new evidence showing how concentrating reward/punishment power in one group member affects cooperation.

Keywords: rewards; punishment; discretionary incentives; decentralized incentives; peer-to-peer incentives; centralized incentives; experiment.

JEL Classification Numbers: C72; H41.

Acknowledgements: We thank Paul van Lange, Michalis Drouvelis, Jan Potters, Abhijit Ramalingam and seminar participants at the University of East Anglia for useful comments. We acknowledge the support of the Leverhulme Trust (ECF/2010/0636) and the Network for Integrated Behavioural Science (ES/K002201/1).

* CeDEx, School of Economics, University of Nottingham, Nottingham, NG7 2RD, United Kingdom.
Nosenzo: email: daniele.nosenzo[at]nottingham.ac.uk, tel. +44 115 846 7492
Sefton: email: martin.sefton[at]nottingham.ac.uk, tel. +44 115 846 6130

1. INTRODUCTION

Social dilemmas stem from the misalignment of individual and group incentives. The optimal decision of a self-interested individual is in conflict with what is best from a societal view. A classic example, and one that we focus on here, is the situation where individuals can voluntarily contribute to public good provision. While all members of a group benefit when an individual contributes to the public good, each individual in the group has a private incentive to free-ride. Standard economic theory, based on the assumption that individuals maximize their own payoff, predicts under-provision relative to the social optimum. There is a long tradition in economics of studying mechanisms that may improve matters, for example by introducing externally-imposed incentives to encourage contributions and discourage free-riding such as subsidies to contributors or taxes on free-riders. Chen (2008) describes many of these mechanisms and reviews related experimental research.

An important alternative approach relies on self-governance (Ostrom, 1990). Here, rather than relying on externally-imposed incentives, groups may design institutional arrangements that let individual group members set and enforce their own norms of cooperation, by voluntarily rewarding fellow group members who contribute and/or by punishing those who free-ride. Most of the literature in economics has focused on arrangements that involve decentralized incentive systems whereby all group members can monitor and reward/punish each other. However, in many settings the power to reward or punish is not distributed equally across all group members, and is often concentrated in the hands of a central monitor. This raises a natural question of how the distribution of reward and punishment power affects their success in promoting cooperation.

The focus of this article is to address this question. To do this, in Section 2 we survey the existing experimental economics literature on decentralized and centralized incentive systems. We focus on discretionary incentives, where group members can voluntarily reward or punish others as opposed to externally-imposed incentives, where group members react to institutionalized rewards and/or punishments.[1] In Section 3, we report a new experiment that examines the relative success of decentralized and centralized rewards and punishments in

[1] Examples of institutionalized incentives are studied in Andreoni and Gee (2012); Croson et al. (2007); Dickinson and Isaac (1998); Dickinson (2001); Falkinger et al. (2000); Fatás et al. (2010). Also related is Yamagishi (1986), where players can contribute to a centralized 'sanctioning fund' which is then used to mete out sanctions on low contributors (see also Sigmund et al., 2010).

sustaining cooperation. Section 4 offers some concluding comments on the broader implications of these findings.

2. THE USE OF INCENTIVES TO PROMOTE COOPERATION IN PUBLIC GOODS GAMES

One of the most extensively used frameworks for the study of cooperation in groups is the 'public goods game' (PGG, henceforth). There are n players, each endowed with E tokens. Each player chooses how many tokens to place in a 'private account' and how many to place in a 'group account'. A player receives a monetary payoff of α from each token placed in her private account. Each token a player contributes to the group account generates a return nfl, which is equally shared among all n members of the group. Thus, from each token she contributes to the group account a player receives a monetary payoff of fl. The game is parameterized such that players have a private incentive to allocate all tokens to private accounts ($fl < \alpha$), whereas the group as a whole would maximize monetary payoffs if all players fully contributed to the group account ($nfl > \alpha$). Thus, tokens placed in the group account are akin to contributions to a public good, and this game captures the tension between private and collective interests which lies at the heart of social dilemmas.[2]

A large number of economic experiments have studied behavior in the PGG (see Chaudhuri, 2011; Ledyard, 1995 for reviews of the literature). A stylized finding is that, although individuals do make positive contributions to the public good, contribution levels typically fall substantially short of the socially efficient level. Moreover, contributions tend to decline with repetition, and full free-riding often prevails towards the end of an experiment. Thus, incentives to free-ride undermine cooperation in PGGs, and the design of institutional arrangements that induce individuals to eschew narrow self-interest and promote cooperation is thus an important issue, which has attracted ubiquitous interest among behavioral scientists.

In this section we review two such arrangements: the use of sanctions against free-riders vis-à-vis the use of rewards for cooperators. In particular, we will review evidence from economic experiments on the effectiveness of punishment and reward incentives when these are administered through a decentralized system, whereby each group member monitors and

[2] This framework in which tokens are allocated between private and group accounts was introduced by Marwell and Ames (1979); Isaac et al. (1984) modified their design to introduce the version described above.

sanctions/rewards other group members, or through a centralized system, whereby the power to administer the incentives is concentrated in the hands of a restricted number of group members.

Standard theory predicts that the opportunity to punish or reward other group members will have no effect on cooperation: costly punishment/reward reduces earnings and so should not be used by a self-interested individual. Knowing this, individuals should not be deterred from pursuing their selfish interests by the threat of punishment, or the promise of rewards. Nevertheless, as we discuss in the next sub-section, the availability of punishments and/or rewards can promote cooperation and increase public good provision relative to settings where discretionary incentives are unavailable.

2.1 Decentralized (Peer-to-Peer) Punishments and Rewards

Most of the economics literature has focused on decentralized (peer-to-peer) incentives, which can be used by all group members to reward or punish each other. The vast majority of studies have focused on punishment incentives (e.g., Fehr and Gächter, 2000; Fehr and Gächter, 2002).[3] Fehr and Gächter (2000), for example, use a two-stage PGG. In the first stage, players choose a contribution level as in the standard game described earlier. In the second stage players learn the contributions of the other members of their group and then simultaneously decide whether to assign 'punishment tokens' to each group member. Each assigned token is costly to both the punishing and the punished player. The availability of peer-to-peer punishment is found to significantly increase contributions relative to the standard PGG: averaging across all periods and treatments reported in Fehr and Gächter (2000), players contribute about 25% of their endowment in the game without punishment, and 67% in the game with punishment.[4] Moreover, the availability of punishment incentives stabilizes cooperation: in the games with punishment contributions do not decrease as in the standard game without punishment, and can increase over time to converge to nearly full cooperation. These findings have been widely replicated in the literature (for recent surveys see, e.g., Chaudhuri, 2011, Gächter and Herrmann, 2009; Shinada and Yamagishi, 2008), and suggest that decentralized punishment systems can provide powerful incentives to cooperate in social dilemmas.

[3] There are parallel literatures focusing on punishment incentives in other related social dilemmas, such as common pool resource dilemmas (Ostrom et al., 1992) or prisoner's dilemmas (Caldwell, 1976).

[4] Punishment opportunities increase contributions both in a "partner" treatment, where players are matched with the same other group members repeatedly, and a "stranger" treatment, where players are re-matched into new groups after each period, though contributions are lower in the latter treatment.

The effectiveness of punishment incentives has also been shown to vary with the punishment technology. In particular, the use of a punishment token imposes costs on both punisher and punishee, and cooperation rates are generally higher and more stable with higher impact-to-cost ratios (e.g., Egas and Riedl, 2008; Nikiforakis and Normann, 2008). 'Low-power' punishments with a 1:1 impact-to-cost ratio are not always successful in encouraging cooperation in PGGs. For example, while Masclet and Villeval (2008) and Sutter et al. (2010) find that 1:1 punishments significantly increase contributions relative to a standard PGG with no punishment, Egas and Riedl (2008), Nikiforakis and Normann (2008) and Sefton et al. (2007) find that 1:1 punishments are ineffective.

In contrast to the abundant evidence on the effectiveness of punishments, the use of rewards to promote cooperation has received less attention in the experimental literature. Sefton et al. (2007) studied a two-stage PGG where, in the first stage, players choose a contribution to the public good and, in the second stage, after having observed others' contributions, players can assign 'reward tokens' to each other. Each token assigned is costly to the rewarding player (her earnings decrease by \$0.10), and increases the earnings of the rewarded player (also by \$0.10). Sefton et al. (2007) find that the availability of rewards increases contributions relative to a standard PGG with no rewards or punishments, although the effect is small (subjects contribute 43% of their endowment in the standard PGG and 59% in the game with rewards) and is statistically significant only at the 10% level. Moreover, rewards do not stabilize cooperation: in the last period of the game with rewards contributions are actually lower (albeit not significantly so) than in the standard game with no punishment/reward incentives.

Other studies have confirmed that when rewards are pure monetary transfers between players, as in Sefton et al. they have only a weak (and mostly statistically insignificant) impact on cooperation rates in PGG experiments (Drouvelis and Jamison, 2012; Sutter et al., 2010; Walker and Halloran, 2004). However, if the impact of the reward on the rewarded player's payoff exceeds the cost of using the instrument, rewards have been found to be effective in encouraging cooperation (Rand et al., 2009 and Sutter et al., 2010, both using 'high-power'

rewards with a 3:1 impact-to-cost ratio). Moreover, high-power rewards are as effective as high-power punishments (Drouvelis, 2010; Rand et al., 2009; Sutter et al., 2010).[5]

These findings point to a potential limitation on the use of rewards to encourage cooperation in social dilemmas: rewards can be expected to be effective only when the cost of assigning the reward is outweighed by the benefits that accrue to the recipient of the reward. While there may be situations where recipient's valuation of the reward exceeds the cost of delivering it (e.g., the awards and perks used in employee recognition programs usually impose modest costs on the firm but may have special value to employees), in many settings rewards that generate direct net benefits may be unavailable.

In summary, the findings in the literature suggest that both peer-to-peer punishments and rewards can effectively promote cooperation in social dilemmas.[6] Crucial to the effectiveness of either instrument is the ratio of the benefit/cost of receiving the reward/punishment to the cost of delivering it. High-power rewards and punishments are both beneficial for cooperation. For a given effect on cooperation, rewards have an efficiency-advantage over the punishment instrument since the mere use of high-power rewards raises joint payoffs (e.g. Rand et al., 2009; Sutter et al., 2010). In contrast, punishment can enhance efficiency only if the efficiency gains from higher contributions exceed the social loss associated with the use of the instrument (see, e.g., Ambrus and Greiner, 2012; Gächter et al., 2008; Herrmann et al., 2008). Thus, while the instruments may be similarly effective in raising contribution levels, (high-power) rewards may be preferred to sanctions on efficiency grounds. On the other hand, there is mixed evidence that low-power punishments are effective in raising contributions in PGG experiments, and low-power rewards have largely been found to be ineffective.

[5] For a review of the relative effectiveness of peer rewards and peer punishment see Milinski and Rockenbach (2012). See Balliet et al. (2011) for further discussion and for a meta-analysis of the effectiveness of discretionary and non-discretionary incentives.

[6] Some studies have examined whether the joint availability of punishments and rewards can further increase cooperation. Results are mixed. Sefton et al. (2007) find that contributions are highest when group members can use both instruments. Rand et al. (2009) find that combining punishments and rewards does not lead to higher contributions than when only punishment or only rewards can be used. Finally, in Drouvelis and Jamison (2012) contributions are higher when both instruments are available than when only rewards are available, but the joint availability of punishments and rewards does not increase contributions relative to a treatment where only punishment is available (however, note that in their experiment the punishment instrument displays a 3:1 impact-to-cost ratio, while the reward instrument has a 1:1 ratio).

2.2 Centralized Punishments and Rewards

While the literature reviewed above suggests that peer-to-peer incentives can successfully promote cooperation, there are also settings where they fail to do so. One problematic aspect of the use of peer-to-peer rewards and punishments is that is that some players may misuse incentives and actually use them to undermine cooperation. For instance, several experiments with peer-to-peer punishment have documented the existence of 'antisocial' or 'perverse' punishment whereby sanctions are used against contributors rather than free-riders with detrimental effects on cooperation (see, for example, Herrmann et al., 2008, Gächter and Herrmann, 2009, or Gächter and Herrmann, 2011).[7] A second issue concerns the potential inefficiencies that may arise if individuals fail to coordinate their punishment or rewarding activities such that too much (or too little) punishment and/or rewarding are meted out. This may be particularly problematic in the case of an excessive use of punishment, due to its associated social inefficiencies. The coordination problems may be further aggravated by the existence of a (second-order) free-rider problem in the use of incentives: since punishing and/or rewarding others is costly, each individual would prefer that someone else bears the burden of enforcing the norm of cooperation (Elster, 1989; Fehr and Gächter, 2002).

Perhaps as a consequence of these difficulties, some groups and societies have developed centralized systems to administer incentives. In such systems, the role of disciplining group members is delegated to one or more authority figures ('monitors') that have exclusive use of the punishment and/or reward instruments. Ostrom (1990), for example, discusses the case of the Hirano, Nagaike and Yamanoka villages in Japan, where the monitoring and sanctioning functions were delegated to 'detectives' who patrolled the communally owned lands and collected fines from villagers caught using the lands without authorization. In some villages, the role of 'detective' was taken up by all eligible male villagers on a rotating basis.[8] Pirate societies are another example of self-organizing groups facing social dilemmas (e.g. in the provision of effort during a plunder) who delegated the administration of discipline to a central authority. The

[7] Several studies have examined whether further punishment stages, in which players can punish punishing behaviors, may be a way to discipline perverse or anti-social punishment. Cinyabuguma et al. (2006) find that allowing such "second-order punishment" has little effect in deterring perverse punishment, and in fact they observe another form of perverse second-order punishment where the punishers of free-riders are punished. Moreover, Denant-Boemont et al. (2007) and Nikiforakis (2008) find that allowing second-order punishment can even have a negative effect on cooperation.

[8] Similar positions of 'guards' have been created by self-organizing groups for the management of irrigation and forestal systems (see Ostrom, 1990; Ostrom, 1999).

power " to allocate provisions, select and distribute loot (...) and adjudicate crew member conßicts/administer discipline" was transferred from pirate crews into the hands of elected quartermasters, who, together with ship captains, acted as 'social leaders' in pirate societies (Leeson, 2007, p. 1065; see also Leeson, 2009). A further, more contemporary example of centralized incentive systems can be found in the arrangement of team incentives in organizations, where the role of administering reward and punishment incentives to team members is concentrated in the hands of a team leader or supervisor.

Delegation of the disciplining power to a central monitor can successfully solve some of the issues of peer-to-peer incentives outlined above. For example, centralizing the use of incentives may eliminate or reduce the inefficiencies arising from the miscoordination of punishing/rewarding activities. The second-order free-riding problem may also be mitigated in the sense that, although the use of incentives is still costly, monitors know that they cannot free-ride on others' monitoring efforts. Moreover, the effectiveness of the punishment and reward instruments may increase when the use of the instruments is centralized. For instance, a monitor who can make a coordinated use of the group resources that are earmarked for the disciplining activity may be able to inflict a harsher punishment upon a free-rider than if the sanctioning resources are diffused across group members. Similarly, the disciplining effect of rewards may increase if these are concentrated in the hands of a monitor who has discretion to allocate or withhold the funds from group members.

On the other hand, there are also potential disadvantages associated with centralized incentive systems. Installing a central monitor to administer the group punishing/rewarding resources does not solve the issue whether the incentives will be used in the best interest of the group. In fact, monitors may face stronger incentives than group members to abuse their power. For example, monitors may be tempted to underuse the resources earmarked for the monitoring and disciplining activities if they are residual claimants of such resources. Analogously, monitors may be tempted to overuse the resources at their disposal, e.g. if they can keep a share of the fines collected from the sanctioned individuals. This discussion highlights the importance of keeping monitors accountable to the group for their activities (Ostrom, 1990). A further issue, discussed in Balliet et al. (2011), regards the relation between monitors and group members. If monitors are perceived as 'out-groups', who are extraneous to the group, this may undermine group cohesion with negative effects on cooperation.

Overall, the discussion above raises interesting questions about the relative effectiveness of centralized incentive systems vis-à-vis decentralized systems in social dilemma settings. Surprisingly, very few experimental studies in economics have focused on centralized punishment and reward systems.[9] One such study is van der Heijden et al. (2009), who examine a team production setting where team members automatically receive an equal share of the team output irrespective of their contribution to team production. This setting is analogous to the public goods setting, and complete free-riding is the unique equilibrium. Van der Heijden et al. (2009) compare this setting to one where the distribution of team output is not automatic, but instead is administered by a 'team leader'. The team leader can monitor other team members' contributions and decide how to allocate team output amongst team members. In this setting team leaders face the temptation to keep the whole team output for themselves, and thus it is unclear whether they can successfully promote cooperation. In fact, van der Heijden et al. (2009) find that the introduction of a team leader significantly increases average team contributions by 73%, and team earnings by 37%. However, not all teams perform well under a team leader: cooperation breaks down when team leaders abuse their power and distribute output unfairly among team members.

Another study focusing on centralized incentives is Gürerk et al. (2009). They study a team production game where, after all team members have made a contribution to team production and received an equal share of team output, one team member (the team leader) receives a monetary budget that she can use to discipline team members. Team leaders can choose the type of incentives that they will have available in the game: they can choose to discipline team members using either punishments or rewards. Gürerk et al. (2009) find that team leaders are initially reluctant to choose the punishment instrument, and almost all team leaders commit to using the reward instrument instead. However, rewards are less effective than punishments in encouraging team members to contribute: contributions decline over time when rewards are used, whereas there is an increasing trend in contributions under punishment incentives. As a consequence, leaders' preference for rewards tends to diminish over time.[10]

[9] Eckel et al. (2010) study a "star-network" setting where one player occupies a central position, but that player can be punished as well as punish. Thus, although punishment power is asymmetrically distributed across group members, it is not centralized in the hands of one player. Similarly, Leibbrandt et al. (2012) and Nikiforakis et al. (2010) study punishment mechanisms with asymmetric (but not centralized) distributions of the power to punish.
[10] A similar finding is reported in Gürerk et al. (2006) for decentralized incentive systems.

These two studies suggest that centralized incentive systems can successfully promote cooperation in social dilemmas. Nevertheless, some of the empirical findings (e.g., the heterogeneity in team success under leaders in van der Heijden et al., 2009) highlight the concerns discussed above about the potential pitfalls of centralized systems. Moreover, since neither study includes treatments where subjects can mutually monitor and punish/reward each other, it is difficult to draw conclusions about the relative success of centralized systems vis-à-vis decentralized systems.

Two recent studies which do include a comparison between centralized and decentralized incentive systems are O'Gorman et al. (2009) and Carpenter et al. (2012). Both studies focus on punishment incentives. O'Gorman et al. (2009) study a PGG with two types of sanctioning systems: one treatment uses the standard mutual monitoring system (peer-to-peer punishment), whereas the other treatment uses a centralized monitoring system whereby only one group member (randomly selected each period) can punish group members. O'Gorman et al. (2009) find that in a treatment without punishment contributions display the usual decreasing trend over time, and average contributions are significantly lower than in either punishment treatment. Contributions do not differ significantly between the centralized and peer-to-peer punishment treatments, but average earnings are significantly higher with centralized punishment than with peer-to-peer punishment, suggesting that a more coordinated use of punishment power can reduce efficiency losses.

Carpenter et al. (2012) also study a PGG and compare a treatment with peer-to-peer punishment and a treatment where the monitoring and punishment power is concentrated in the hands of one group member. In contrast to O'Gorman et al. (2009), Carpenter et al. (2012) report higher contributions under peer-to-peer punishment (where subjects on average contribute 56% of their endowment) than in the treatment with centralized punishment (where average contributions are 33% of the endowment). Average earnings are also higher under peer-to-peer than centralized punishment. There are several differences in the experimental designs of the O'Gorman et al. (2009) and Carpenter et al. (2012), but perhaps a potential explanation for the different findings reported in the two studies is that in O'Gorman et al. (2009) the role of central monitor was randomly assigned to a new subject in each new round of play, whereas in Carpenter et al. (2012) roles remained constant throughout the experiment. Assigning the role of central monitor on a rotating base, as in O'Gorman et al. (2009), may attenuate the negative

impact that the appointment of a 'bad monitor' (e.g., a monitor who never sanctions free-riding) may have on contribution dynamics, and may thus explain the different success of the centralized system across the two studies.[11]

While the studies reviewed above shed some light on the effectiveness of centralized incentives systems in sustaining cooperation in social dilemmas, several important questions remain unanswered. First, while the studies by O'Gorman et al. (2009) and Carpenter et al. (2012) allow a comparison between centralized and decentralized punishment systems, we are not aware of any study comparing centralized and decentralized reward systems. Second, with the exception of van der Heijden et al. (2009), all other studies have examined 'high-power' incentives (with either a 3:1 or 2:1 impact-to-cost ratio). As discussed in the previous sub-section, 'high-power' peer-to-peer punishments and rewards are usually very successful in encouraging cooperation in social dilemmas, and so it seems that there may be little scope for centralized systems to improve over this. An interesting question is whether concentrating punishment/reward power can improve the success of incentives in environments where peer-to-peer incentives have been less successful in encouraging cooperation. In this sense, the result reported in van der Heijden et al. (2009) that concentrating rewards in the hands of a leader is beneficial for cooperation is interesting because most of the literature on 1:1 peer-to-peer rewards find them to be ineffective, and so this suggests that concentrating reward power may increase the effectiveness of the instrument. In the next section we report a new experiment to shed light on some of these questions.

3. THE EFFECTIVENESS OF CENTRALIZED VS. DECENTRALIZED INCENTIVES: NEW EVIDENCE

3.1 Design and Procedures

The new experiment used 150 volunteer participants, randomly matched into three-person groups.[12] All groups played a two-stage PGG, repeated over ten periods in fixed groups. In stage

[11] Baldassarri and Grossman (2011) study a rather different sort of centralized punishment institution where a third-party (i.e. a player external to the group who neither contributes nor benefits from contributions) is responsible for punishment. They find that centralized punishment is most effective where the third-party is elected rather than randomly appointed. Relatedly, Grechenig et al. (2012) study a centralized punishment institution where a player external to the group controls the allocation of group members' punishment resources. Their focus, however, is on group members' choices between this centralized-punishment institution and standard decentralized and no-punishment institutions. In the setting most similar to those studied in O'Gorman et al. (2009) and Carpenter et al. (2012), they find that subjects prefer decentralized punishment to centralized and no-punishment institutions.

one each player received an endowment of 20 tokens and had to choose how many to allocate to a public account and how many to keep in a private account. A player earned 3 points for each token she kept in her private account, and 2 points from each token allocated to the public account (regardless of which group member had contributed it). At the end of the stage players were informed of the decisions and earnings of each group member. In stage two 24 additional tokens were given to each group and these could be used to either increase own-earnings, or reward or punish other group members, depending on treatment. At the end of stage two players were informed of the stage two decisions and earnings of each group member. Players were then informed of their total earnings for the period, these being the sum of their earnings from the two stages.

Altogether we had five treatments, with ten groups in each treatment. In our Baseline treatment 8 stage-two tokens were placed in each player's private account, yielding 3 points per token. In our mutual monitoring treatments each player was given 8 stage-two tokens that could be kept in a private account, yielding 3 points per token, or assigned to other group members. In our Mutual-Punish treatment a token assigned to a group member reduced her earnings by 3 points, while in our Mutual-Reward treatment each stage two token assigned to a group member increased her earnings by three points. Based on the previous research discussed earlier we expected that the effect of allowing 1:1 rewards or punishment would be quite weak, allowing for the possibility that concentrating reward/punishment power would increase the effectiveness of rewards or punishment.

In our central monitoring treatments all punishment/reward power was concentrated in the hands of one, randomly assigned, group member (this was the same group member in all ten periods). This group member (the 'central monitor') was given 24 tokens in stage two, and these could be placed in her private account, yielding 3 points per token, or assigned to other group members. Assigning a token reduced the assignee's earnings by three points in the Central-Punish treatment, and increased her earnings by three points in the Central-Reward treatment.[13]

---

[12] The experiment was conducted in multiple sessions in the CeDEx computerized laboratory at the University of Nottingham in Spring 2012. Subjects were recruited using ORSEE (Greiner, 2004) and the software was written in z-tree (Fischbacher, 2007). A copy of experimental materials is available in the Appendix.

[13] In addition to earnings from each period we gave each participant an initial point balance of 320 points (which covered any potential losses in the treatments allowing punishment). At the end of a session the initial balance plus accumulated point earnings from all periods was converted into cash at a rate of £0.75 per point. Participants earned, on average, £11.10 for a session lasting between 30 and 60 minutes.

Our parameters were chosen to satisfy two criteria that we considered may be important for the successful administration of incentives. First, we chose parameters to give central monitors the ability to discipline. A group member can increase her stage one earnings by 20 points by keeping all her tokens rather than contributing them. We wanted to give the central monitor sufficiently many stage two tokens so that she could reduce a free-rider's stage two earnings by at least 20 points by withholding rewards in the Central-Reward treatment. With 24 stage two tokens a monitor could employ a strategy of rewarding full contributors with 8 tokens and withholding rewards from free-riders. Against this strategy a defector gains 20 points in stage one and loses 24 points in stage two, and so this strategy effectively disciplines free-riders. Second, we wanted to give central monitors an incentive to discipline. Since the central monitor could free-ride in stage one and keep all her stage two tokens, giving a payoff of at least (20 x 3) + (24 x 3) = 132 points, we wanted the payoff from full contribution and an equal allocation of reward tokens to be higher than this. With our parameters the central monitor receives (60 x 2) + (8 x 3) = 144 points when the group contributes fully and she rewards equally.

3.2 Results

As seen in Figure 1, our Baseline treatment exhibits the standard pattern observed in previous public goods experiments. Average contributions start out around 60% of endowments and steadily decline with repetition to 20% of endowments in period ten. Across all rounds participants contribute, on average, 43% of their endowments. Relative to this, we find that peer punishment is highly effective. Averaging across all periods participants contribute 89% of their endowments in the Mutual-Punish treatment. This difference in average contributions is highly significant ($p = 0.001$).[14] In contrast, peer rewards are ineffective: average contributions in the Mutual-Reward treatment are 45% of endowments, not significantly different from Baseline ($p = 0.821$). These findings are qualitatively consistent with the findings from exogenous 1:1 treatments reported in Sutter et al. (2010). However, punishment is much more effective in our study, perhaps reflecting the stronger punishment technology.[15]

[14] All p-values are based on two-sided Wilcoxon rank-sum tests treating each group as a unit of observation.
[15] In their study the gain to a player from free-riding completely rather than contributing fully was 8 experimental currency units, whereas the most punishment that a player could receive was 3 experimental currency units.

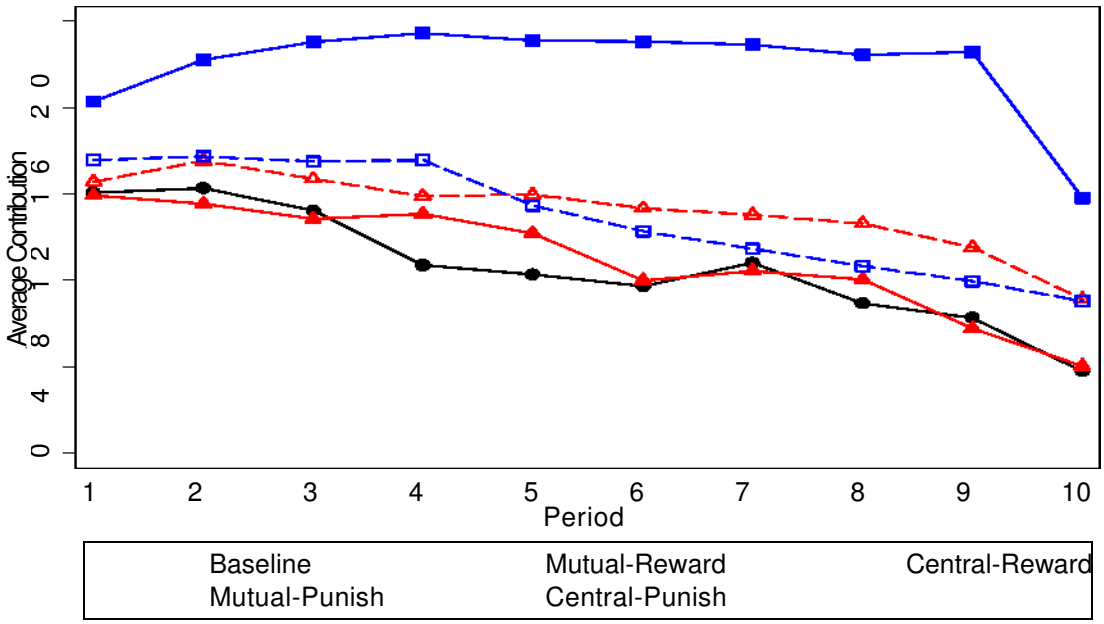


Figure 1: Average contributions to public good across periods.

Turning to our Central-Punish treatment, we find that concentrating punishment power reduces the effectiveness of punishment. Contributions are 55% of endowments in Central-Punish, significantly lower than in Mutual-Punish ($p = 0.038$). In fact, when punishment is left to a central monitor average contributions are not significantly different from Baseline ($p = 0.571$). Concentrating reward power, on the other hand, results in a small, but insignificant, increase in contributions. Average contributions in Central-Reward are 56% of endowments, not significantly different from Mutual-Reward ($p = 0.406$) or Baseline ($p = 0.406$).

Differences between the Mutual-Punish treatment and the other treatments are also evident in Figure 2, which shows the distribution of contributions, pooling over all periods. While in the Mutual-Punish treatment only 5% of decisions were to contribute zero, in all other treatments around 30% of contribution decisions result in extreme free-riding. At the other extreme, in the Mutual-Punish treatment players contribute 20 tokens 78% of the time, twice as often as in any other treatment. In turn, full contributions are twice as frequent in the centralized monitoring treatments as in Baseline (39% in Central-Punish, 38% in Central-Reward, and 18% in Baseline).

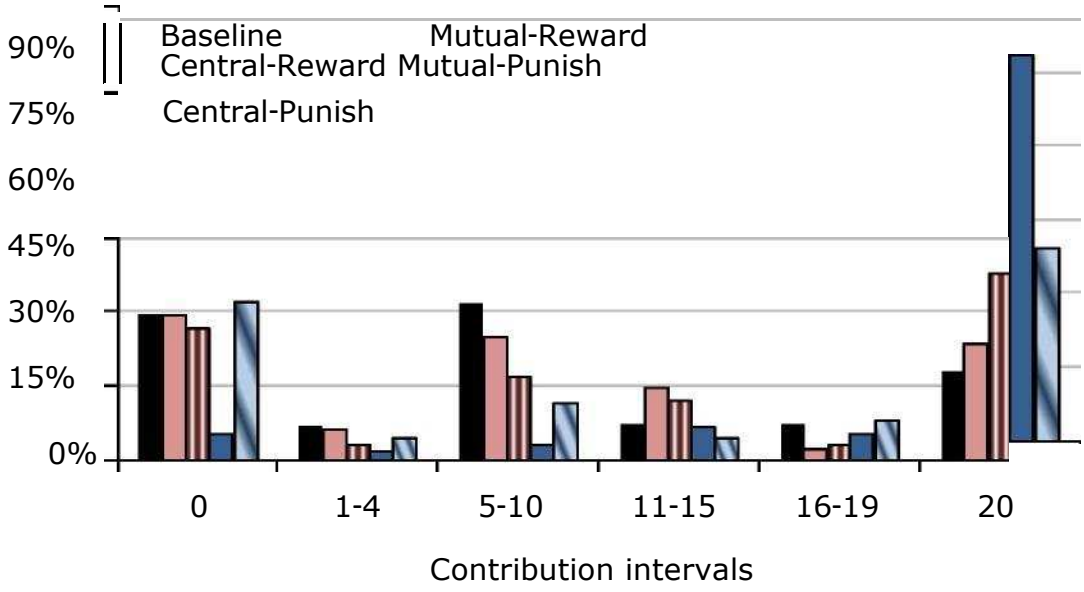


Figure 2. Distribution of individual contributions to public good.

Differences between treatments seem most strong for the case of maximal contributions. While the proportion of maximal contributions is significantly higher in Mutual-Punish than Baseline ($p = 0.001$), the differences between proportions in the centralized monitoring treatments and Baseline are not significant (Central-Reward versus Baseline, $p = 0.255$; Central-Punish versus Baseline, $p = 0.128$). The reason for this is not so much the size of the effect (though as Figure 2 shows, there is a considerable difference in this respect), but rather the reliability of the effect. In nine of ten Mutual-Punish groups maximal contributions are observed as often as not, while the same can be said of only one Baseline group. In the centralized monitoring treatments there is more heterogeneity. In each treatment there are four groups that look like Mutual-Punish groups in terms of their propensity to contribute fully, and six groups that look like Baseline groups.

This heterogeneity across groups in contribution behavior translates into heterogeneity in earnings. In principle, a group could earn as much as 144 points per group member per period, and in fact two groups in Mutual-Punish achieved this, but there was considerable variability in actual earnings among the other groups. Figure 3 presents box-and-whiskers diagrams for the

distribution of group earnings, where the box shows the lower quartile, median, and upper quartile of attained group earnings.[16]

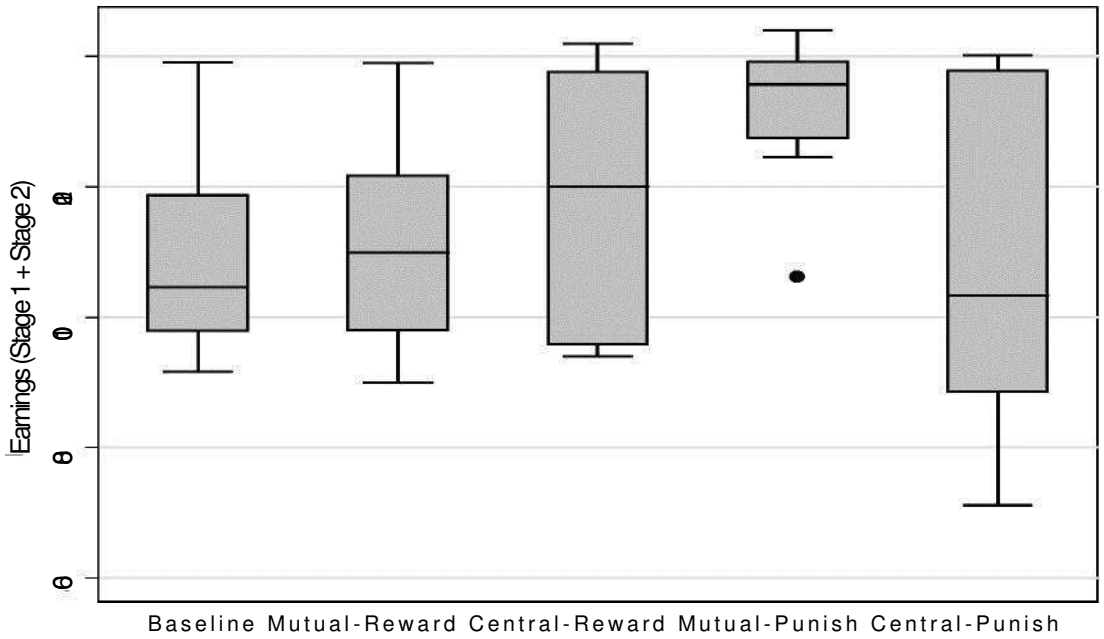


Figure 3. Box plots of group performance. Group performance is measured by earnings per group member per period.

Relative to Baseline, earnings are higher and less variable in the Mutual-Punish treatment. It is worth noting that even in the short horizon of the experiment, earnings are significantly higher in Mutual-Punish than Baseline ($p = 0.008$). The diagrams also show that centralized monitoring leads to more variable group performances than the other treatments. Apparently, when placed in the role of central monitor some players used the strategy of contributing fully and disciplining free-riders, and successfully managed a cooperative group. At the same time, other central monitors failed either because they used punishments or rewards ineffectively, or because they did not try to use them at all. This result underlines the importance of differences in leadership qualities across individuals. This was previously noted by van der Heijden et al. (2009), who also observed a mixture of 'good' and 'bad' leaders. Overall leadership seemed much more effective in their study, although the power technology of the leaders is very different across studies. An interesting avenue for further research could be to identify the psychological mechanisms underlying differences in leadership quality.

[16] The whiskers represent the lowest (highest) observation still within 1.5 times the inter-quartile range of the lower (upper) quartile. One outlier is indicated in the Punish treatment.

Next we examine how rewards and punishments were directed. Figure 4 shows that the number of reward/punishment tokens a subject receives is sensitive to how her contribution relates to the average contribution of the rest of her group.

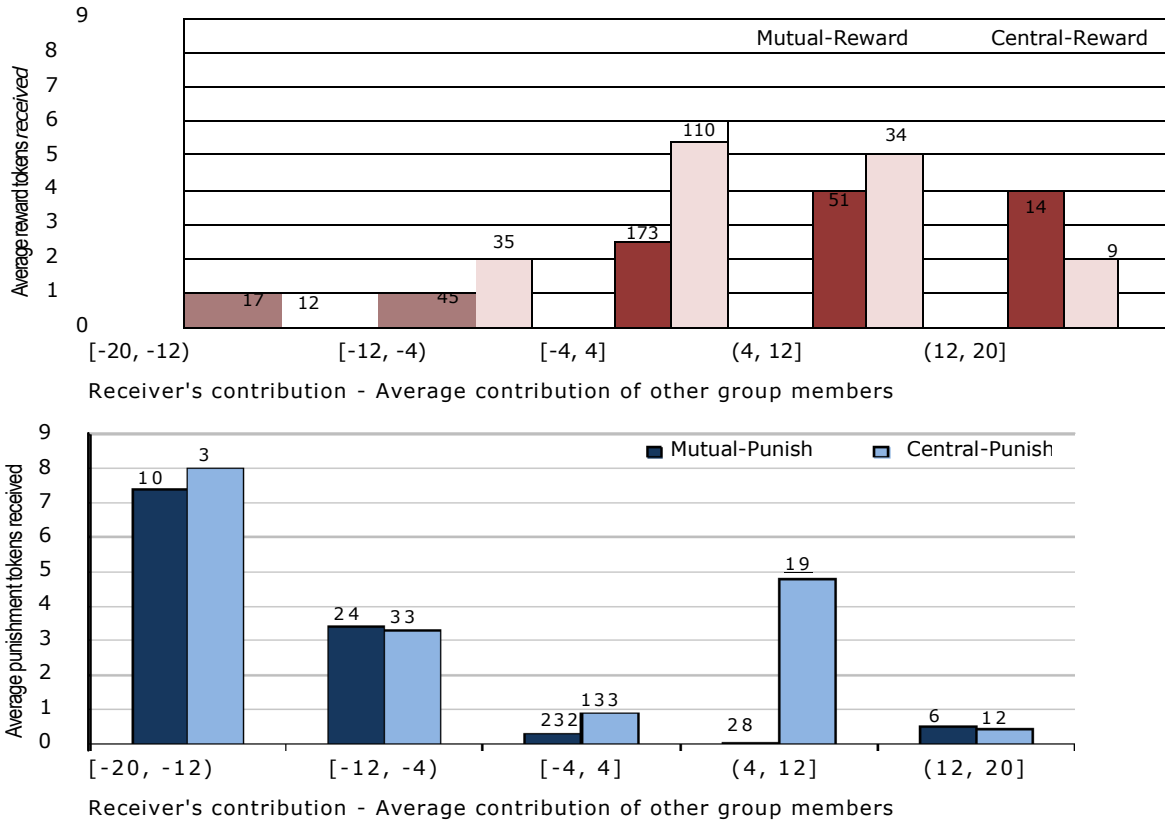


Figure 4. Reward/punishment tokens received. Numbers above bars indicate the number of cases in each interval.

The patterns in the mutual monitoring treatments are similar to previous experiments. Group members are punished more heavily the lower is their contribution relative to the average contributed by the rest of their group, and rewards are mainly targeted at those who contribute at or above the average contributed by the rest. However, large deviations below the average of others' contributions are punished more heavily than cooperators are rewarded. On average, if a player free-rides rather than contributes fully in a group where the others contribute fully, she receives more than 20 points of punishment in Mutual-Punish but forgoes far less than 20 points of rewards in Mutual-Reward. There is a noticeable difference in the way incentives are used in the Central-Punish treatment, where punishments are more frequently meted out against group

members who contribute above the group average. Such anti-social punishment has been shown to be detrimental to cooperation in mutual monitoring environments (see section 2 above).[17]

The effectiveness of punishment and rewards depends crucially on how players respond to the use of these incentives. Figure 5 shows how individuals change their contributions in response to punishment/reward received in the previous period. We distinguish between those subjects who give less than, the same as, and more than, the group average. In the punishment treatments those who contribute less than the group average tend to increase their contribution in response to being punished. In Mutual-Punish there is very little punishment of those who contribute at or above the average of the rest of their group, but in Central-Punish these players are sometimes punished and they tend to respond by decreasing their contribution. In the reward treatments those that contribute less than the average of the rest of their group do not change their contribution much, while those who contribute at or above the average of the rest of the group tend to decrease their contribution if they don't get rewarded. This asymmetry in the response to rewards may explain why they have little impact in sustaining cooperation.

In summary, our new experiment finds that peer-to-peer punishment is an effective mechanism for promoting cooperation, whereas peer-to-peer rewards are much less effective. Our treatments with concentrated reward/punishment power were designed so that central monitors had both an incentive to induce cooperation, and sufficient resources to be able to incentivize cooperation, by other group members. Nevertheless, cooperation in these treatments is not significantly higher than in our Baseline.

[17] Although some potential explanations for anti-social punishment cannot operate in a centralized monitoring environment (e.g. revenge for being punished in previous periods, or pre-emptive retaliation by free-riders who expect to get punished), some others can (e.g. a dislike of non-conformism). It is difficult to explain why anti-social punishment is more prevalent with centralized as opposed to mutual monitoring. One possibility is that players can discipline anti-social punishers in the mutual monitoring environment by punishing them in subsequent periods, while the central monitor does not have to fear punishment. If so, then this suggests that the possibility to discipline central monitors (for example by voting them out of office) might reduce anti-social punishment by central monitors. We chose to appoint central monitors randomly for purposes of experimental control, but future research could examine how the effectiveness of central monitors depends on the way they are appointed.

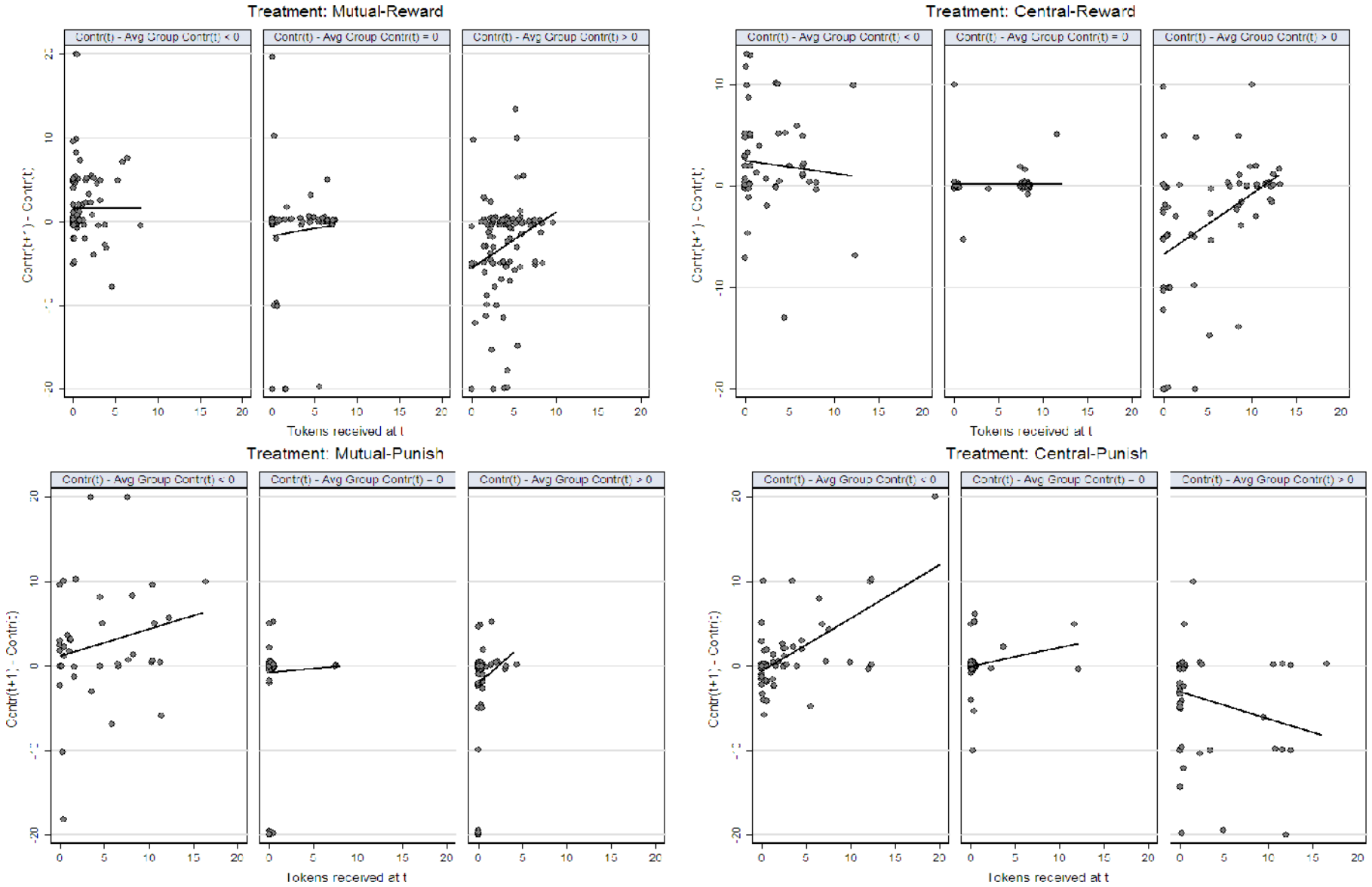


Figure 5. Response to reward/punishment tokens received. For each treatment we show three separate OLS regression lines, one based on data from players who contributed less than the average of the other members of their group (leftmost panel), one for players who contributed the same as the average of the other group members (middle panel), and one for players who contributed more than the average of the other group members (rightmost panel).

4. CONCLUSIONS

We reviewed evidence on the effectiveness of discretionary incentives to promote cooperation. A large literature in economics shows that decentralized (peer-to-peer) punishments can be effective in raising contributions and earnings. Previous studies have emphasized the importance of high-power incentives and/or long time horizons for the success of punishment institutions. Our new experiment shows that punishments can be effective even within a context of a 'low power' incentive system (with a 1:1 impact-to-cost ratio) and short horizon (10 periods). Indeed, somewhat surprisingly, peer-to-peer punishment elicited (and maintained) high levels of cooperation from very early in the game, indicating that the threat of punishment seems almost instantly recognized by our subjects. Peer-to-peer rewards, on the other hand, have not been studied as extensively. While high-power rewards are as effective as high-power punishments, most studies find that low-power rewards are ineffective. Our new experiment with low-power rewards confirms this.

Our primary focus, however, is to compare these decentralized incentive systems with centralized systems. Centralized systems have natural manifestations in the organization of teams and small groups and societies, where the administration of incentive power is delegated to subsets of group members (which we refer to as 'central monitors'). Perhaps surprisingly, centralized systems have been relatively overlooked in the literature. Existing evidence suggests that concentrating incentive power may enhance efficiency. However, some of the findings caution that conferring all reward or punishment power on one individual is risky: some individuals are found to abuse (or at least, fail to use effectively) their incentive power, with detrimental consequences for cooperation. Our new experiment finds that concentrating punishment power clearly reduces its effectiveness, while concentrating reward power slightly raises its effectiveness, albeit not significantly.

The poor performance of centralized institutions in our setting may partly reflect the specific details of the decision environment studied in our experiment. For example, for centralized institutions to be successful central monitors must have a sufficient incentive to use rather than keep for themselves the group resources earmarked for the disciplining activity. This "incentive to discipline" is sensitive to the parameters of the decision environment. For example, in our setting the incentive to discipline increases with the size of the group, ceteris paribus.

Thus, centralized institutions may be found to be more effective in settings where central monitors control larger groups. Another condition for centralized institutions to be successful is that central monitors have sufficient disciplining power to induce a player who only cares about her own material payoff to contribute. This "ability to discipline" is also sensitive to the parameters of the decision setting. For example, in our experiment the ability to discipline increases in the amount of resources earmarked for the disciplining activity. Thus, central monitors may be more successful in settings where they control a larger amount of resources.

Our new experiment also finds considerable heterogeneity in group performance when power is centralized: while some groups perform well under a central monitor, in other groups central monitors simply keep to themselves resources that can be used to discipline others. To the extent that central monitors advance their own material payoff by keeping these resources and not using them to incentivize cooperation through the disciplining of free-riders, this can be viewed as an abuse of their power. This highlights the importance of designing appropriate constraints to the ability of monitors to abuse their power. Centralized systems should not be transformed into autocratic systems where monitors cannot be held accountable for their use of group resources. On the contrary, centralized systems should have built-in mechanisms to allow group members to oversee the conduct of those who are empowered with disciplining authority.

Another source of failure in our central monitoring treatments comes from monitors who use the resources erratically and thus fail to establish a clear norm of cooperation. This creates uncertainty about whether actions will be punished and/or rewarded and this may undermine the usefulness of incentives. Such uncertainty could be reduced if monitors could use other devices (such as communication) to induce norms of cooperation and explain the likely consequences of adhering or deviating from these norms. Nevertheless, we suspect that some monitors would still fail even with unlimited opportunities to communicate their intentions.

This points to the importance of selecting the right individuals for the role of monitor. Thus, how monitors are appointed is an important feature of centralized systems. For example, systems where the monitoring role is assigned to group members on a rotating basis may limit the negative impact of 'bad monitors', and thus have an advantage over systems with consolidated power positions. The appointment of central monitors through democratic elections (as it happened for captains and quartermasters in pirate societies) may be another mechanism that allows group members to screen out candidate monitors who are likely to abuse their power.

REFERENCES

Ambrus, A., and B. Greiner. 2012. Imperfect Public Monitoring with Costly Punishment - An Experimental Study. American Economic Review 102(7), 3317–32.

Andreoni, J., and L.K. Gee. 2012. Gun for hire: Delegated enforcement and peer punishment in public goods provision. Journal of Public Economics 96(11–12), 1036–1046.

Baldassarri, D., and G. Grossman. 2011. Centralized sanctioning and legitimate authority promote cooperation in humans. Proceedings of the National Academy of Sciences.

Balliet, D., L.B. Mulder, and P.A.M. Van Lange. 2011. Reward, punishment, and cooperation: A meta-analysis. Psychological Bulletin 137(4), 594–615.

Caldwell, M.D. 1976. Communication and sex effects in a five-person Prisoner's Dilemma Game. Journal of Personality and Social Psychology 33(3), 273–280.

Carpenter, J., S. Kariv, and A. Schotter. 2012. Network architecture, cooperation and punishment in public good experiments. Review of Economic Design 16(2-3), 93–118.

Chaudhuri, A. 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics 14(1), 47–83.

Chen, Y. 2008. Incentive-compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research. In Handbook of Experimental Economics Results, Volume 1:625–643. Elsevier.

Cinyabuguma, M., T. Page, and L. Putterman. 2006. Can second-order punishment deter perverse punishment? Experimental Economics 9(3), 265–279.

Croson, R., E. Fatás, and T. Neugebauer. 2007. Excludability and contribution: a laboratory study in team production. CBEES Working paper 07-09, University of Texas, Dallas.

Denant-Boemont, L., D. Masclet, and C.N. Noussair. 2007. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Economic Theory 33(1), 145–167.

Dickinson, D.L. 2001. The carrot vs. the stick in work team motivation. Experimental Economics 4(1), 107–124.

Dickinson, D.L., and R.M. Isaac. 1998. Absolute and relative rewards for individuals in team production. Managerial and Decision Economics 19(4-5), 299–310.

Drouvelis, M. 2010. The Behavioural Consequences of Unfair Punishment. University of Birmingham Department of Economics Discussion Paper 10-34.

Drouvelis, M., and J. Jamison. 2012. Selecting Public Goods Institutions: Who Likes to Punish and Reward? Federal Reserve Bank of Boston, WP No. 12-5.

Eckel, C.C., E. Fatás, and R. Wilson. 2010. Cooperation and Status in Organizations. Journal of Public Economic Theory 12(4), 737–762.

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

Elster, J. 1989. The cement of society. A study of social order. Cambridge: Cambridge University Press.

Falkinger, J., E. Fehr, S. Gächter, and R. Winter-Ebmer. 2000. A simple mechanism for the efficient provision of public goods: Experimental evidence. American Economic Review 90(1), 247–264.

Fatás, E., A.J. Morales, and P. Ubeda. 2010. Blind justice: An experimental analysis of random punishment in team production. Journal of Economic Psychology 31(3), 358–373.

Fehr, E., and S. Gächter. 2000. Cooperation and punishment in public goods experiments. American Economic Review 90(4), 980–994.
Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. Nature 415(6868), 137–140.
Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171–178.
Gächter, S., and B. Herrmann. 2009. Reciprocity, culture, and human cooperation: Previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B – Biological Sciences 364(1518), 791–806.
Gächter, S., and B. Herrmann. 2011. The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. European Economic Review 55(2), 193–210.
Gächter, S., E. Renner, and M. Sefton. 2008. The long-run benefits of punishment. Science 322, 1510.
Grechenig, K., A. Nicklisch, and C. Thöni. 2012. Information-sensitive Leviathans: The Emergence of Centralized Punishment. Mimeo, University of Lausanne.
Greiner, B. 2004. An Online Recruitment System for Economic Experiments. In Forschung und wissenschaftliches Rechnen. GWDG Bericht 63, ed. K. Kremer and V. Macho, 79–93. Göttingen: Ges. für Wiss. Datenverarbeitung.
Gürerk, Ö., B. Irlenbusch, and B. Rockenbach. 2006. The competitive advantage of sanctioning institutions. Science 312(108), 108–111.
Gürerk, Ö., B. Irlenbusch, and B. Rockenbach. 2009. Motivating teammates: The leader's choice between positive and negative incentives. Journal of Economic Psychology 30(4), 591–607.
Heijden, E. van der, J. Potters, and M. Sefton. 2009. Hierarchy and opportunism in teams. Journal of Economic Behavior & Organization 69(1), 39–50.
Herrmann, B., C. Thöni, and S. Gächter. 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Isaac, R.M., J.M. Walker, and S.H. Thomas. 1984. Divergent Evidence on Free Riding - an Experimental Examination of Possible Explanations. Public Choice 43(2), 113–149.
Ledyard, J.O. 1995. Public goods: A survey of experimental research. In The Handbook of Experimental Economics, ed. Alvin E. Roth and John H. Kagel, 111–181. Princeton: Princeton University Press.
Leeson, P.T. 2007. An-arrgh-chy: The Law and Economics of Pirate Organization. Journal of Political Economy 115(6), 1049–1094.
Leeson, P.T. 2009. The calculus of piratical consent: the myth of the myth of social contract. Public Choice 139(3), 443–459.
Leibbrandt, A., A. Ramalingam, L. Sääksvuori, and J.M. Walker. 2012. Broken Punishment Networks in Public Goods Games: Experimental Evidence. Jena Economic Research Paper 2012-004.
Marwell, G., and R. Ames. 1979. Experiments on the provision of public goods I: Resources, interest, group size, and the free-rider problem. American Journal of Sociology 84(6), 1335–1360.
Masclet, D., and M.C. Villeval. 2008. Punishment, inequality, and welfare: a public good experiment. Social Choice and Welfare 31(3), 475–502.
Milinski, M., and B. Rockenbach. 2012. On the interaction of the stick and the carrot in social dilemmas. Journal of Theoretical Biology 299, 139–143.

Nikiforakis, N. 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves? Journal of Public Economics 92(1-2), 91–112.
Nikiforakis, N., and H. Normann. 2008. A comparative statics analysis of punishment in public goods experiments. Experimental Economics 11(4), 358–369.
Nikiforakis, N., H.-T. Normann, and B. Wallace. 2010. Asymmetric Enforcement of Cooperation in a Social Dilemma. Southern Economic Journal 76(3), 638–659.
O'Gorman, R., J. Henrich, and M. van Vugt. 2009. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. Proceedings of the Royal Society B-Biological Sciences 276(1655), 323–329.
Ostrom, E. 1990. Governing the commons: The evolution of institutions for collective action, the political economy of institutions and decisions. Cambridge: Cambridge University Press.
Ostrom, E. 1999. Coping with Tragedies of the Commons. Annual Review of Political Science 2(1), 493–535.
Ostrom, E., J.M. Walker, and R. Gardner. 1992. Covenants with and without a sword - Self-governance is possible. American Political Science Review 86(2), 404–417.
Rand, D.G., A. Dreber, T. Ellingsen, D. Fudenberg, and M.A. Nowak. 2009. Positive Interactions Promote Public Cooperation. Science 325(5945), 1272–1275.
Sefton, M., R. Shupp, and J.M. Walker. 2007. The effect of rewards and sanctions in provision of public goods. Economic Inquiry 45(4), 671–690.
Shinada, M., and T. Yamagishi. 2008. Bringing Back Leviathan into Social Dilemmas. In New Issues and Paradigms in Research on Social Dilemmas, ed. Anders Biel, Daniel Eek, Tommy Gärling, and Mathias Gustafsson, 93–123. Springer US.
Sigmund, K., H.D. Silva, A. Traulsen, and C. Hauert. 2010. Social learning promotes institutions for governing the commons. Nature 466(7308), 861–863.
Sutter, M., S. Haigner, and M.G. Kocher. 2010. Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. Review of Economic Studies 77(4), 1540–1566.
Walker, J.M., and M.A. Halloran. 2004. Rewards and sanctions and the provision of public goods in one-shot settings. Experimental Economics 7(3), 235–247.
Yamagishi, T. 1986. The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology 51(1), 110–116.

APPENDIX – Experimental Instructions

[ALL TREATMENTS]

Instructions

Welcome!

You are about to participate in a decision-making experiment. Please do not talk to any of the other participants until the experiment is over. If you have a question at any time please raise your hand and an experimenter will come to your desk to answer it.

At the beginning of the experiment you will be matched with two other people, randomly selected from the participants in this room, to form a group of three. **The composition of your group will stay the same throughout the experiment**, i.e. you will form a group with the same two other participants during the whole experiment. Each person in the group will be randomly assigned a role, either 'group member A', 'group member B' or 'group member C'. **Your role will stay the same throughout the experiment**. Your earnings will depend on the decisions made within your group, as described below. Your earnings will not be affected by decisions made in other groups. All decisions are made anonymously and you will not learn the identity of the other participants in your group. You will be identified simply as 'group member A', 'group member B' and 'group member C'.

At the beginning of the experiment you will be informed of your role and given an initial balance of 320 points. The **experiment will then consist of 10 periods**, and in each period you can earn additional points. At the end of the experiment each participant's initial balance plus accumulated point earnings from all periods will be converted into cash at the exchange rate of 0.75 pence per point. Each participant will be paid in cash and in private.

Description of a period

Every period has the same structure and has two stages.

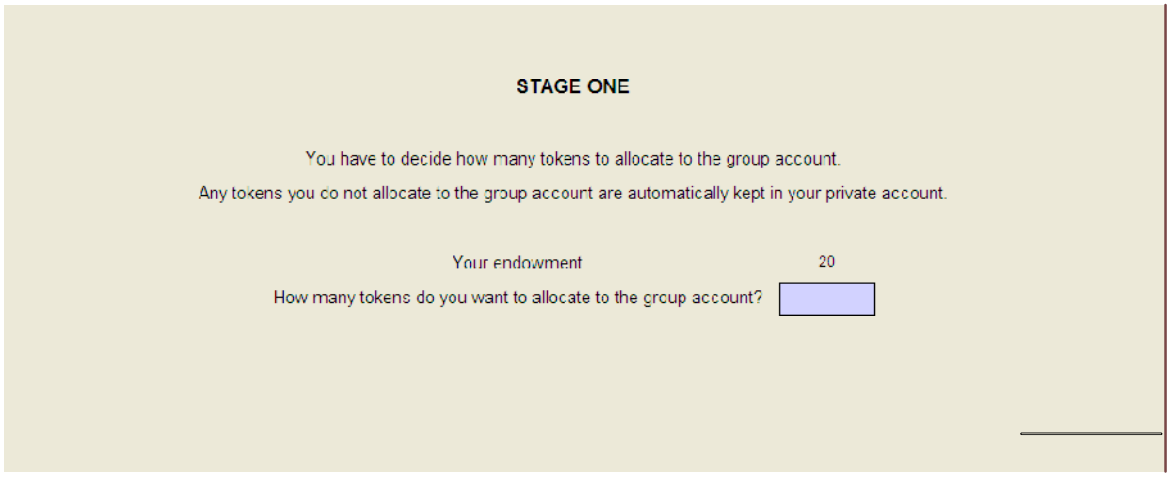
Stage One

In Stage One of each period you will be endowed with 20 tokens.

You must choose how many of these tokens to allocate to a group account and how many to keep in your private account.

Similarly, the other two members of your group will be endowed with 20 tokens each and must choose how many tokens to allocate to the group account and how many to keep in their private accounts.

You will make your decision by entering the number of tokens you allocate to the group account. Any tokens you do not allocate to the group account will automatically be kept in your private account. You enter your decisions on a screen like the one shown below.



Earnings from Stage One will be determined as follows:

For each token you keep in your private account you will earn 3 points.

For each token you allocate to the group account you and the other two members of your group will earn 2 points each.

Similarly, for each token another group member keeps in his or her private account this group member will earn 3 points, and for each token he or she allocates to the group account all three group members will earn 2 points each.

Your point earnings from Stage One will be the sum of your earnings from your private account and the group account.

Thus:

> **Your point earnings from Stage One = 3 x (number of tokens kept in your private account) + 2 x (total number of tokens allocated to the group account by yourself and the other two members of your group).**

Before we describe Stage Two we want to check that each participant understands how earnings from Stage One will be calculated. To do this we ask you to answer the questions below. In a couple of minutes the experimenter will check your answers. When each participant has answered all questions correctly we will continue with the instructions.

Stage One Questions

1. How many periods will there be in the experiment? ____

2. How many people are in your group (including yourself)? ____

3. Will you be matched with the same or different people in every period? (circle one)

SAME DIFFERENT

4. Suppose in Stage One of a period each group member allocates 0 tokens to the group account.

How many tokens does A keep in his or her private account? _____

What will be A's earnings from his or her private account?

What is the total number of tokens allocated to the group account?

What will be A's earnings from the group account?

What will be A's earnings from Stage One?

5. Suppose in Stage One of a period each group member allocates 20 tokens to the group account.

How many tokens does A keep in his or her private account? _____

What will be A's earnings from his or her private account? _____

What is the total number of tokens allocated to the group account? _____

What will be A's earnings from the group account? _____

What will be A's earnings from Stage One? _____

6. Suppose A allocates 16 tokens to the group account, B allocates 10 tokens to the group account, and C allocates 4 tokens to the group account.

How many tokens does A keep in his or her private account?

What will be A's earnings from his or her private account?

What is the total number of tokens allocated to the group

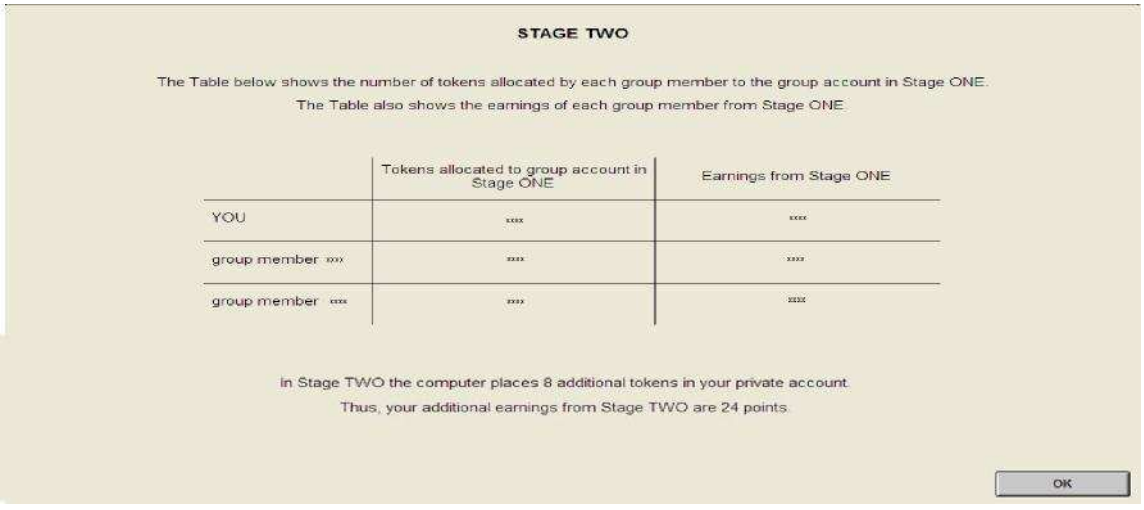
account? What will be A's earnings from the group account?

What will be A's earnings from Stage One?

[BASELINE]

Stage Two

At the beginning of Stage Two you will be informed of the decisions made by each group member and their earnings from Stage One in a screen like the one below.



In Stage Two of each period you will be endowed with 8 additional tokens. These will automatically be placed in your private account from which you earn 3 points per token. Thus you will earn an additional 24 points in Stage Two.

Similarly, the other group members will be endowed with 8 additional tokens each, which will be automatically placed in their private accounts from which they earn 3 points per token. Thus the other group members will earn an additional 24 points each in Stage Two.

Neither you nor the other group members make any decisions in Stage Two.

Ending the period

At the end of Stage Two the computer will inform you of your total earnings for the period. Your period earnings will be the sum of your earnings from Stage One and Stage Two.

At the end of period 10 your accumulated point earnings from all periods will be added to your initial balance of 320 points to give your total point earnings for the experiment. These will be converted into cash at the exchange rate of 0.75 pence per point. Each participant will be paid in cash and in private.

[MUTUAL-PUNISH / MUTUAL-REWARD]

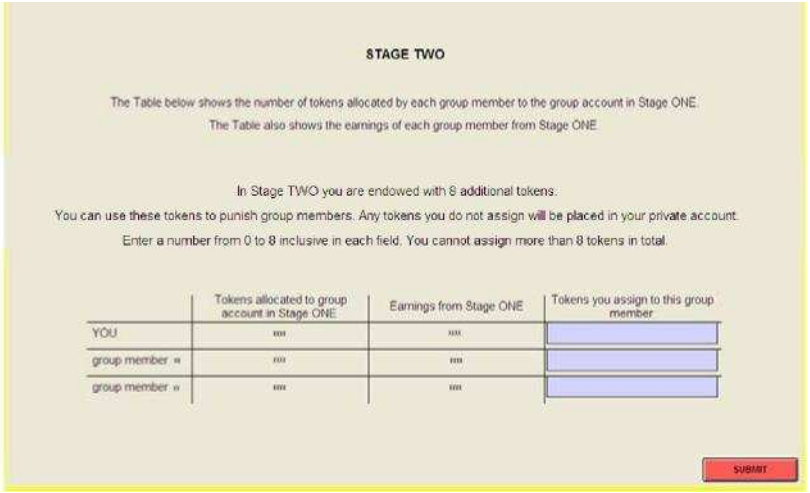
Stage Two

At the beginning of Stage Two you will be informed of the decisions made by each group member and their earnings from Stage One.

In Stage Two of each period you will be endowed with 8 additional tokens. You must choose how many of these to use to [punish] [reward] group members and how many to keep in your private account.

Similarly, the other group members will be endowed with 8 additional tokens each and must choose how many of these to use to [punish] [reward] group members and how many to keep in their private accounts.

You make your decision by completing a screen like the one below. You choose how many tokens to assign to each group member. You can assign tokens to yourself if you want. Any of the 8 additional tokens not assigned will automatically be kept in your private account. You cannot assign more than 8 tokens in total.



Earnings from Stage Two will be determined as follows:

For each additional token you keep in your private account you will earn 3 points.

For each token you assign to a group member that group member's earnings will be [reduced] [increased] by 3 points.

Similarly, for each additional token another group member keeps in his or her private account this group member will earn 3 points, and for each token he or she assigns to a group member that group member's earnings will be [decreased] [increased] by 3 points.

Thus:

> **Your point earnings from Stage Two = 3 x (number of additional tokens kept in your private account) [-] [+] 3 x (total number of tokens assigned to you by all group members in Stage Two).**

We want to check that each participant understands how their earnings from Stage Two will be calculated. To do this we ask you to answer the questions below. In a couple of minutes the experimenter will check your answers. When each participant has answered all questions correctly we will continue with the instructions.

Stage Two Questions

Suppose in Stage Two of a period A assigns 2 tokens to B and 2 tokens to C. B assigns 2 tokens to B and 6 tokens to C. C assigns 8 tokens to B.

1. How many tokens does A keep in his or her private account? _____
2. What is the total number of tokens assigned to A? _____
3. What will be group member A's earnings from Stage Two? _____
4. How many tokens does B keep in his or her private account? _____
5. What is the total number of tokens assigned to B? _____
6. What will be group member B's earnings from Stage Two? _____
7. How many tokens does C keep in his or her private account? _____
8. What is the total number of tokens assigned to C? _____
9. What will be group member C's earnings from Stage Two? _____

Ending the period

At the end of Stage Two the computer will inform you of all decisions made in Stage Two and the earnings of each member of your group for Stage Two. The computer will then inform you of your total earnings for the period. Your period earnings will be the sum of your earnings from Stage One and Stage Two.

At the end of period 10 your accumulated point earnings from all periods will be added to your initial balance of 320 points to give your total point earnings for the experiment. These will be converted into cash at the exchange rate of 0.75 pence per point. [Although your earnings in some periods may be negative, your initial balance ensures that your final earnings for the experiment cannot be negative.] Each participant will be paid in cash and in private.

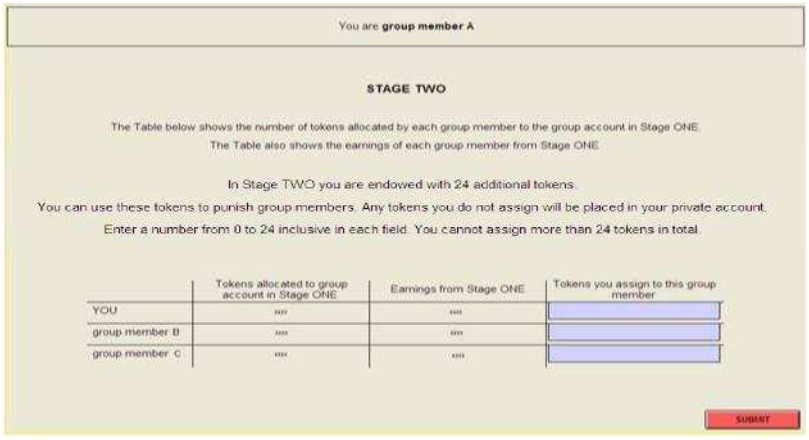
[CENTRAL-PUNISH / CENTRAL -REWARD]

Stage Two

At the beginning of Stage Two you will be informed of the decisions made by each group member and their earnings from Stage One.

In Stage Two of each period group member A will be endowed with 24 additional tokens. Group member A must choose how many of these to use to [punish] [reward] group members and how many to keep in his or her private account. Group members B and C do not make any decisions in Stage Two.

If you are group member A, you make your decision by completing a screen like the one below. You choose how many tokens to assign to each group member. You can assign tokens to yourself if you want. Any of the 24 additional tokens not assigned will automatically be kept in your private account. You cannot assign more than 24 tokens in total.



Earnings from Stage Two will be determined as follows:

For each additional token A keeps in his or her private account he or she will earn 3 points.

For each token A assigns to a group member that group member's earnings will be [decreased] [increased] by 3 points.

Thus:

> **A's point earnings from Stage Two = 3 x (number of additional tokens kept in A's private account) [–] [+] 3 x (number of additional tokens A assigns to himself or herself)**
>
> **B's point earnings from Stage Two = [–] [+] 3 x (number of additional tokens assigned to B by group member A).**
>
> **C's point earnings from Stage Two = [–] [+] 3 x (number of additional tokens assigned to C by group member A).**

We want to check that each participant understands how their earnings from Stage Two will be calculated. To do this we ask you to answer the questions below. In a couple of minutes the experimenter will check your answers. When each participant has answered all questions correctly we will continue with the instructions.

Stage Two Questions

Suppose in Stage Two of a period A assigns 12 tokens to B and 8 tokens to C.

1. How many tokens does A keep in his or her private account?

2. What is the total number of tokens assigned to A?
3. What will be group member A's earnings from Stage Two?
4. What is the total number of tokens assigned to B?
5. What will be group member B's earnings from Stage Two?
6. What is the total number of tokens assigned to C?
7. What will be group member C's earnings from Stage Two?

Ending the period

At the end of Stage Two the computer will inform you of all decisions made in Stage Two and the earnings of each member of your group for Stage Two. The computer will then inform you of your total earnings for the period. Your period earnings will be the sum of your earnings from Stage One and Stage Two.

At the end of period 10 your accumulated point earnings from all periods will be added to your initial balance of 320 points to give your total point earnings for the experiment. These will be converted into cash at the exchange rate of 0.75 pence per point. [Although your earnings in some periods may be negative, your initial balance ensures that your final earnings for the experiment cannot be negative.] Each participant will be paid in cash and in private.

[ALL TREATMENTS]

Beginning the experiment

If you have any questions please raise your hand and an experimenter will come to your desk to answer it.

We are now ready to begin the decision-making part of the experiment. Please look at your computer screen and begin making your decisions.