Davide Barrera, Vincent Buskens and Vera de Rover

18 Comparing Consequences of Carrots and Sticks on Cooperation in Repeated Public Good Games

Abstract: Many sociologists and economists have maintained that costly sanctions are able to sustain cooperation, but whether carrots or sticks are more successful in this respect is still under dispute (e.g., Balliet, Mulder, and Van Lange 2011; Rand et al. 2009; Sefton, Schupp, and Walker 2007). Furthermore, while many studies investigated the effects of sanctioning institution on cooperation, the long-term effects of sanctions on group solidarity are largely unexplored. In this chapter, we discuss contrasting hypotheses concerning the effects of positive and negative sanctions on cooperation in Public Good Games and solidarity among the group members. Subsequently, we test these hypotheses by means of a laboratory experiment. Our results show that while carrots do increase cooperation, sticks turn out to be more effective. Concerning group solidarity, we do not find differences in group solidarity depending on the type of sanctions available to the group members. However, we find that actors who receive rewards show higher solidarity towards the group.

18.1 Introduction

In many practical instances, individual interests are partly conflicting with collective interests. For example, everyone prefers to receive public services from the state, but, at the same time, everyone would prefer not to pay taxes that allow the state to provide those services. As it is individually rational to avoid paying taxes, but collectively detrimental if everyone does so, we call this a social dilemma (Raub, Buskens, and Corten 2015). Real-life situations with this kind of incentive structure are modelled using public good games (PGG hereafter). The typical setup of a PGG is as follows. People interact in small groups. Each individual receives an initial endowment and can decide how much of this endowment he or she wants to invest into a 'group account,' and how much he or she wants to keep for him- or herself. In a PGG, the amount contributed by all group members to the group account grows by a certain rate and – thereafter – is equally distributed among all the group members, regardless of their contribution to the public good. The growth rate of the public good is always lower than the number of people in the group. This

Davide Barrera, University of Turin and Collegio Carlo Alberto
Vincent Buskens, Vera de Rover, Department of Sociology / ICS, Utrecht University

https://doi.org/10.1515/9783110647495-018

implies that, for every unit an actor invests in the public good, he or she receives only a fraction in return. Therefore, it is individually rational not to invest in the group account.

Although standard game-theoretical rationality assumptions dictate that nobody should cooperate in public good games, some cooperation is always observed when subjects play public good games in experimental laboratories (see Chauduri 2011 and Ledyard 1995 for reviews). Several mechanisms accounting for the emergence of cooperation have been proposed, such as kin selection (Hamilton 1963), direct reciprocity (Trivers 1971), social embeddedness (Buskens and Raub 2013; Raub and Weesie 1990), indirect reciprocity (Nowak and Sigmund 1998), group selection (Traulsen and Nowak 2006), status (Willer 2009). See Nowak (2006) for an overview. In the context of cooperation in PGGs, a large body of literature focused on the role of institutions. Analyzing the problem of social order, Hobbes (1651) described an authority who enforces rules for the society by establishing a contract. Social order obtains because violations of this contract are punished by the central authority (the "Leviathan" in Hobbes's terms). More generally, institutional solutions to public good problems include all forms of sanctions that can be positive (i.e., rewards or carrots) or negative (i.e., punishments or sticks).[1] If actors are given the possibility to use sanctions, they typically use them as selective incentives to promote cooperation (Olson 1965). The idea that sanctions are an effective means to protect cooperation from free-riders is widely known and has been largely debated (Balliet, Mulder, and van Lange 2011; Barclay 2006; Fehr and Gächter 2002; Fehr and Gintis 2007; Gächter 2013; Yamagishi 1986). While Balliet, Mulder, and Van Lange (2011) convincingly show – using an extensive meta-analysis – that both sticks and carrots can promote cooperation and that the two effects do not seem too much different, the number of studies with costly rewards in PGGs is limited. In addition, the consequences of using sanctions, particularly punishment, have been considered only in terms of their ability to promote cooperation. Yet, the fact that Hobbes (1651) used a biblical monster, the leviathan, as a metaphor for the central authority responsible of maintaining social order, suggests that order imposed by force is not necessarily always pleasant. In fact, some scholars have addressed possible side-effects of punitive sanctions concerning solidarity among the groups members (Fehr and Rockenbach 2003; Irwin, Mulder, and Simpson 2014; Mulder et al. 2006; Nowak et al. 2008; Oliver 1980). In this chapter, we focus on two questions: first, whether carrots or sticks are a more effective way to obtain cooperation in a public good problem, and second, what are the consequences of positive and negative sanctions on how much solidarity individuals feel towards their fellow group members. In the remainder, after briefly discussing the existing relevant literature, we present a series of hypotheses concerning effects of positive and negative

1 See Calvert (1995) on modeling institutions as exogenous constraints.

sanctions on both cooperation and solidarity. Subsequently, we present an experiment in which these hypotheses are empirically tested.

18.2 Theory and hypotheses

18.2.1 Theory: Human cooperation and selective incentives

Before discussing the theoretical arguments related to the use of positive versus negative sanctions, we formally introduce the PGG. In a PGG, groups consist of N (with $N > 2$) persons. Each group member is endowed with a specific amount (Y) of points. The amount of points is the same for every person. Each member can invest a certain amount X ($0 \leq X \leq Y$) of these points into a group account representing the public good. Every point invested in the group account is multiplied by a factor M (which exceeds 1 but is smaller than N, $1 < M < N$). After all group members have made their contribution to the group account and the total amount of points has been multiplied by M, the resulting amount of points is equally distributed among the group members. Because $M < N$, the individual return of the investment is negative, i.e. $X > X \times M/N$. Therefore, choosing $X = 0$ is a dominant strategy for every player. Thus, assuming that actors are purely interested in maximizing their own wealth, they should not contribute anything to the public good. However, if all group members invest their entire endowment Y, each individual earns $N \times Y \times M/N = X \times Y > Y$. This situation is Pareto-superior to the situation in which no one invests anything (Fehr and Gintis 2007; Gächter 2013). In a one-shot PGG, if standard economic rationality is assumed (that is, everyone tries to maximize his expected payoff with no concern for what happens to the other group members), all actors should play the dominant strategy and choose to contribute nothing to the public good. However, this hypothesis is typically not supported by experimental results. In the laboratory there are always some persons who cooperate: in one-shot PGGs, the average contribution level is usually around 50% of the total endowment.

If the PGG is played repeatedly, simple reciprocity (Hamilton 1963; Trivers 1971) signaling (Spence 1973; Zahavi 1995) or reputation (Axelrod 1984; Kreps et al. 1982; Raub and Weesie 1990) become possible mechanisms supporting contribution to the public good. Nevertheless, repeated PGGs typically show a pattern of decreasing contribution over time empirically. In experimental treatments in which the group composition is kept constant (*partner* treatment), the mean contribution is higher than when participants are shuffled after every round (*stranger* treatment), but in both cases contribution declines as the game proceeds further (e.g., Andreoni 1988; Keser and Van Winden 2000). These empirical results imply that reciprocity, reputation and costly signaling are not sufficient to reach a cooperative equilibrium in a repeated PGG. The pattern of declining cooperation is reversed and higher cooperation

levels are achieved and sustained when sanctions are allowed (Fehr and Gächter 2002) – even if participants are shuffled and re-matching is excluded by design. Fehr and Gachter's (2002) study on altruistic punishment shows that 1) a significant proportion of actors has not purely selfish preferences, 2) these actors typically contribute significantly to the public good and punish defectors, 3) as defectors anticipate that they will be punished, they raise their cooperation levels, and, consequently, cooperation is ensured. The existence of (partly) altruistic preferences poses an evolutionary puzzle because selfish actors gain a competitive advantage in interactions with these altruists. Therefore, altruistic preferences should have been driven extinct by natural selection.

Some scholars have argued that these preferences are a form of maladaption (e.g., Johnson, Stopka, and Knights 2003). The argument goes as follows: in the ancestral environment where most of our evolution took place – sometimes referred to as "environment of evolutionary adaptiveness", EEA (e.g., Fehr and Henrich 2003) – interactions were typically *repeated* rather than one-shot, because most of the relevant social activities occurred within small groups. In such settings, reputation-building makes cooperation profitable, because defection might lead to retaliation and conflict, thus making defectors worse off in the long run. As one-shot interactions are assumed to have been rare in EEA, human subjects interacting anonymously in the laboratory fail to recognize the irrelevance of reputation building in one-shot interactions and erroneously apply behavioral rules evolved and adapted to repeated interactions.

However, this argument is problematic because one-shot interactions were in fact not so rare in EEA and, moreover, experimental subjects do show the ability to differentiate their behavior depending on whether interactions are repeated or one-shot (Fehr and Henrich, 2003). To solve this puzzle, scholars have developed culture-gene coevolution models that provide evolutionary foundations for altruistic preferences (e.g., Boyd et al. 2003). In such models actors face cooperation problems within groups and, simultaneously, competition for scarce resources between groups. Consequently, selection favors defectors within groups, but groups where defectors thrive underperform in intergroup competition and risk extinction. In this environment, groups developed some form of institutional solution – such as punishment or reward – to prevent defectors to exploit cooperators thereby producing a reduction of the group collective fitness, ultimately leading to extinction of the group. Thus, these models typically include some form of group selection, but they assume that group selection operates predominantly on cultural rather than genetic variation (Richerson, Boyd and Henrich 2003). Accordingly, altruistic preferences evolved in the EEA lead to the development of various institutions, which in turn are transmitted to subsequent generations through processes like conformist social learning and moralistic enforcement of norms (Richerson, Boyd and Henrich 2003). However, if institutions are assumed to be transmitted via conformist social learning, group solidarity and individual commitment to the group must be important determinants of a group's capacity to survive and

reproduce itself. Yet, the relation between sanctioning institutions and group solidarity has been largely ignored.

In this chapter we study a setting in which both positive and negative sanctions are allowed and we focus on a repeated game with the same partners because earlier results revealed that, in these conditions, a strong effect of sanctioning institutions is to be expected (Balliet, Mulder, and van Lange 2011; Rand et al. 2009; Sefton, Schupp, and Walker 2007). Also, we focus on groups that remain stable over time, because we are interested in possible side-effects of sanctions, particularly with respect to group solidarity, which only make sense if the group members experience more than just one interaction with each other.

18.2.2 Punishment and rewards

While reciprocity can produce cooperation in two-person games (Axelrod 1984), when the social dilemma involves a group of actors, as in the PGG, reciprocity alone is insufficient, because the individual actions are not *selective*. That is, when the payoff of an actor is affected by more than one person, it is not possible to apply direct reciprocity. In a PGG, either one cooperates with all other group members or with none. Therefore, cooperation is unstable, even in repeated games with the same partners (Andreoni 1988; Keser and Van Winden 2000). The option to impose sanctions, whether in the form of carrots (rewards) or sticks (punishments) addresses precisely this problem of *selecting* incentives to discriminate cooperators from free-riders. By being able to address only the defectors with sticks, or only the contributors with carrots, cooperation can possibly be sustained (Andreoni, Harbaugh, and Vesterlund 2003; Barclay 2006; Fehr and Gächter 2002; Masclet et al. 2003; Rand et al. 2009; Sefton, Schupp, and Walker 2007; Stoop, Van Soest, and Vyrastekova 2018; Yamagishi 1986).

Sanctions are common in everyday life: even just looking at someone angrily or smiling at someone in response to some action could be seen as a stick or carrot, respectively (cf. Masclet et al. 2003; Pan and Houser 2017; Takacs and Janky 2007). In experimental settings, punitive monetary sanctions are often implemented by allowing everyone to impose a (costly) fine on anyone else. For example, each member of a specific group can decrease or increase each other group members' payoffs by paying a small amount from their own earnings. Similarly, rewards can be implemented by allowing everyone to increase the payoff of anybody else, paying a small cost.

According to standard game theoretical assumptions, selective sanctioning institutions imply an additional problem as they likewise require cooperation to be produced. This problem is commonly known as the second-order free-rider dilemma (Yamagishi 1986). Although theoretically rational self-interested actors should not use costly sanctions, many experiments have shown that, when given the opportunity, a considerable proportion of actors invest resources to punish defectors or reward

cooperators. In order to account for this empirical evidence, alternative assumptions concerning the preferences of the actors have been proposed. e.g., *strong reciprocity* (Fehr and Gintis 2007; Gintis 2000). Individuals are assumed to be motivated by strong reciprocity if they cooperate when they expect others to cooperate, reciprocate when they are rewarded and punish violations of the cooperative norm (Fehr and Gintis 2007; Keser and Van Winden 2000). Most commonly, cooperators (people who invest) punish defectors (people who do not invest into the group account) and reward other cooperators (Fehr and Gächter 2002; Gächter 2013; Sefton, Schupp, and Walker 2007).[2] Consequently, as defectors anticipate that they will be punished or that they will not receive any reward, they raise their contribution levels (Fehr and Gächter 2002). Accordingly, the classical model of economic rationality – *homo economicus* – does not accurately account for human behavior in real life. An alternative model, incorporating assumptions of strong reciprocity, is commonly referred to as the PBC-model (Preferences, Beliefs and Constraints- model) (Fehr and Gintis 2007; Gächter 2013). This model considers *P*references, *B*eliefs and *C*onstraints, on the basis of which actors make their decisions. "*Preferences* describe how an individual ranks the available alternatives according to his or her subjective tastes" (Gächter 2013: 34) and these can include more aspects than just his or her own money. That is, partly altruistic preferences are also possible. *Beliefs* refer to what actors think the preferences of other actors might be or how these preferences are distributed. Finally, *Constraints* "describe the set of alternatives that are available to an individual" (Gächter 2013: 34). The PBC-model assumes that individuals choose their utility-maximizing option, an option that satisfies their preferences best, while considering their beliefs and constraints.

18.2.3 Hypotheses on contributions to the public good

In order to derive hypotheses on the effects of monetary sanctions (carrots and sticks) on cooperation in public good type of settings, we assume that actors make their decisions as postulated by the PBC-model. The PBC-model can accommodate various types of preferences, including inequality aversion (Fehr and Schmidt 1999) or fairness (Rabin 1993). Although distinguishing between these alternative non-standard utility models exceeds the scope of our chapter, preferences for fairness and equality are important motivational forces behind the phenomenon of strong reciprocity (Fehr and Gintis 2007; Gächter 2013; Keser and Van Winden 2000). Consistent with the PBC-model, we distinguish two types of individuals: *selfish individuals* (that correspond to the homo economicus) and *conditional cooperators*

2 Antisocial punishment, i.e., punishment directed at high contributors, has been observed especially in cross-cultural experiments (Herrmann, Thöni, and Gächter 2008). However, antisocial punishers seem relatively rare.

(Fischbacher, Gächter, and Fehr 2001).[3] Selfish individuals act according to the standard game theoretical prediction and always try to maximize their own wealth, without any concern for the payoff obtained by others. Conditional cooperators act according to the scheme of strong reciprocity. Assuming a population with heterogeneous preferences solves the second-order free-rider problem, because it can be expected that, in every group, there will be some conditional cooperators whose preference is to punish free-riders and to reward contributors. When sanctions are not allowed, selfish players succeed in driving cooperation down: as conditional cooperators face some defectors, their willingness to contribute declines and finally stops, since defection is the only available means to retaliate against defectors. By contrast, when sanctions are allowed, they are commonly used by conditional cooperators to force selfish actors to contribute (Fehr and Fischbacher 2003; Fehr and Gintis 2007). A symmetric argument can be made for rewards: they will typically be offered by conditional cooperators to high contributors. However, enticed by the possibility to gain rewards, selfish actors will then raise their contribution level, too. Therefore, on the macro level, we state the following hypothesis:

> *H1: Cooperation levels are higher when sanctions (carrots, sticks, or both) can be used than when they are not allowed (baseline).*

When carrots and sticks have to be compared, it is less clear what happens as contrasting arguments can be made favoring both sticks and carrots. Therefore, after summarizing the arguments we will present contrasting hypotheses for the effects of carrots versus sticks on cooperation. The effects of sticks are relevant especially for selfish actors, since they are normally used by conditional cooperators against free-riders. Conditional cooperators are unlikely to be punished, by definition. Conversely, selfish actors are likely to get punished for attempting to free-ride. As defectors get punished, the benefit earned by free-riding is eroded. Therefore, contributing becomes the payoff-maximizing choice. Sticks have thus the effect of increasing contribution of selfish individuals in future rounds of the PGG. On the other hand, carrots are likely to be given especially to conditional cooperators, as they should be the highest contributors. However, carrots affect both conditional cooperators and selfish actors. The former have two motives to contribute: first, they act according to strong reciprocity. Thus, if others cooperate they should cooperate as well. In addition, they may cooperate because they expect to receive rewards for high contributions. As for the selfish actors, carrots create economic advantages: if they are used extensively, they might produce higher payoffs (Rand et al. 2009). Thus, selfish individuals are only motivated to cooperate by the second

3 Fischbacher, Gächter, and Fehr (2001) conducted a study specifically designed to elicit individual preferences. They found that 50% of their participants were conditional cooperators and 30% were selfish types. The remaining 20% displayed other less easily interpretable behavioral patterns.

reason. That is, they will cooperate in order to receive monetary rewards and improve their payoff, if the probability of getting rewarded is high enough.

In addition to the economic argument, there is also a psychological argument favoring the efficacy of carrots over sticks. There is extensive evidence that, next to the positive effect of raising contribution, punitive sanctions produce some pernicious effects: they reduce mutual trust and trustworthiness within groups (Fehr and Rockenbach 2003; Mulder et al. 2006), they increase selfishness and hostility (Nikiforakis 2008), and they crowd out intrinsic motivation (Bohnet and Baytelman 2007). These arguments lead us to present the following hypothesis:

H2a: Cooperation levels are higher when only carrots can be used, compared to when only sticks can be used.

Next to the perspectives that predict advantages of carrots, there are also arguments for advantages of sticks. Dari Mattiacci and De Geest (2009) argue that punishment produces a "multiplication effect": while the sheer *opportunity* to punish is enough to sustain cooperation, punishment may never need to be *actually* used. In other words, assuming that sticks function as credible threats, if everybody cooperates – whether because they are cooperators, or because they are selfish but fear punishment – sticks never need to be applied. An effective stick can provide incentives for the same individual in different periods or for several individuals simultaneously, without anybody ever having to pay its costs. By contrast, carrots need to be used (and paid) each time a subject cooperates. Thus, carrots provide a more expensive means to sustain cooperation.

In addition to this economic argument, there are also psychological arguments favoring sticks over carrots. Individual's utility does not only depend on material benefits: people do not only avoid punishment and seek rewards for financial reasons, but also for the intrinsic benefit that receiving rewards and avoiding punishment provide. According to Lindenberg (2001) and Stigler and Becker (1977), individuals' universal goals include social approval, status, and affection. In particular, as they strive for social approval, people are sensitive to negative and positive sanctions. They strive to receive positive social reactions and avoid negative ones. However, social psychologists have shown that the desire to avoid negative feedbacks is typically stronger than the desire to receive positive feedbacks and this negative asymmetry, due to which "bad is stronger than good" is present across a broad range of phenomena (Baumeister et al. 2001). These arguments lead us to present the following hypothesis:

H2b: Cooperation levels are higher when only sticks can be used, compared to when only carrots can be used.

Since we have no clear prediction on the contribution level when both kinds of sanctions are allowed simultaneously, we turn to the results of Sefton, Schupp, and Walker (2007) and Rand et al. (2009) for comparison. These two studies compared costly rewards and costly punishment and yielded contrasting results, due to

crucial differences in the experimental design. In Sefton, Schupp, and Walker (2007) subjects played series of ten PGGs in groups of four. Each PGG was followed by a second stage in which they received extra points that they could use to punish, reward, and punish or reward (depending on the experimental treatment) the other group members. The ratio between cost and effect of the sanction was 1:1, i.e., one point spent on punishment decreased the income of the person punished of one point and one point spent on reward increased the income of the person rewarded of one point. Because of this feature, in Sefton, Schupp, and Walker (2007) rewards consisted in pure zero-sum horizontal transfer of points, without efficiency gain. In Rand et al.'s (2009) setup the game sequence was the same, but the total number of PGGs to be played was unknown to subjects. In addition, Rand et al. (2009) did not provide the subjects with extra money for the sanctioning part, but sanctions had to be paid from the money earned in the previous PGG. The cost to effect ratio of sanctions was 4:12, i.e., subjects could spend four points on punishment/reward to cause a decrease/increase of twelve points to the target. Finally, unlike in Sefton, Schupp, and Walker (2007), not only the group composition but also the subjects' IDs were held fixed throughout the experiment. Thus, mutual exchange of punishment or rewards – even unrelated to the behavior in the contribution stage of the game – was possible in Rand et al. (2009).

In Sefton, Schupp, and Walker (2007) the treatment in which both sticks and carrots were allowed yielded slightly higher contribution levels than any carrots and sticks separately, while the treatments with rewards only and punishment only produce similar levels of contribution to the public good. By contrast, in Rand et al. (2009) the contribution level was similar in the three treatments, but final payoffs were significantly higher in the two treatments where rewards were possible (i.e., one with only rewards and one with both rewards and punishments) than in the treatment where only punishment was possible. The high efficiency of the treatments with rewards is clearly due to the fact that, as personal IDs of the players were fixed in Rand et al. (2009), actors could exchange rewards mutually and these exchanges were highly productive due to the 4:12 cost to effect ratio. However, the possibility for mutual exchange of rewards effectively transforms the sanctioning stage of the game in a repeated prisoner's dilemma, for which mutual cooperation is a possible equilibrium, irrespectively of what happens in the PGG (Milinski and Rockenbach 2012). In addition, others have shown that, under certain conditions, the possibility of mutual exchanges of rewards could even be detrimental for the production of collective goods, because actors might increase each other's payoffs by exchanging gifts and stop contributing to the public good (Flache 2002).

As detailed in the experimental design section, we adopted Sefton, Schupp, and Walker's (2007) setup, with two modifications: first in order to strengthen the effect of sanctions we increased the magnitude of sticks and carrots by making the cost to effect ratio 1:2 instead of 1:1. Thus, in our study one point spent on sanctions increased or decreased the payoff of the target of two points. Second, we made the series of

rounds longer, in order to allow more time for the effects of sanctions to build up, but we turned the game into an indefinitely repeated game by making the end uncertain for the subjects. Note that we kept the group composition constant throughout the rounds but, unlike Rand et al. (2009), we did only inform subjects about the total number of points all others together used to sanction them, while we did not reveal the identities of these others. In this way, we make mutual exchanges of rewards (or mutual punishment) impossible. Given that our setup largely resembles the one adopted by Sefton, Schupp, and Walker (2007), except for increasing the potential efficacy of sanctions by modifying the cost to effect ratio, we expect to replicate the effect found in their study. Therefore, we propose the following hypothesis:

H3: If carrots and sticks are allowed simultaneously, the cooperation levels are higher than when only either of the two is allowed.

18.2.4 Hypotheses on group solidarity

While previous research has focused primarily on the necessity of sanctioning institutions to promote cooperation in public goods, possible negative side-effects of these institutions have received considerably less attention. However, Tenbrunsel and Messick (1999) argued that the use of monetary sanctions can have perverse effects on group solidarity because it imposes an economic frame, which may crowd out intrinsic motivation and lead subjects to cooperate only to avoid financial losses. The consequences of using carrots and sticks in terms of solidarity among group members have not yet been sufficiently investigated. However, the existence of important side-effects, potentially affecting group solidarity, has been noted by several researchers (Oliver 1980; Fehr and Rockenbach 2003; Mulder et al. 2006, Irwin, Mulder, and Simpson 2014). The existing literature suggests that carrots can promote solidarity and emotional attachment to the group (Markovsky and Lawler 1994; Friedkin 2004). Sticks, on the other hand, may succeed in controlling defection of group members, but could have detrimental side effects on solidarity due to the negative emotional responses they evoke (Oliver 1980; Friedkin 2004). These side effects can range from unhappiness to tension and hostility within the group (Oliver 1980). In general, punishers are disliked while rewarders are liked (Friedkin 2004). Moreover, sticks may crowd out the intrinsic motivation to help the other group members, and damage or undermine mutual trust (Fehr and Rockenbach 2003; Mulder et al. 2006). More specifically, being punished by another group member decreases trust and willingness to cooperate with the punisher (Oliver 1980; Fehr and Rockenbach 2003; Mulder et al. 2006).

As stated above, gene-culture coevolution models of altruistic preferences attribute an important role to cultural group selection in the process leading to the evolution of complex sanctioning institutions. Therefore, intrinsic motivation to support

the group as well as mutual trust and solidarity between group members are very important factors in the competition between groups. Low levels of mutual trust and solidarity might lead groups to perform poorly in competition with other groups. Demotivated group members might be tempted to leave the group. Ultimately, in the long-term a group characterized by an unpopular sanctioning institution might be defeated by other groups or simply dissolve.

The same argument holds for modern groups whose members stay the same for a long time, as it is the case for many small groups (e.g. clubs and organizations). If the members do not feel attached to the group or dislike each other, they will ultimately be less willing to contribute, because feelings have the ability to weaken or strengthen the bonds between group members (Markovsky and Lawler 1994). Although investigating long-term effects in the artificial setting of a computer-mediated laboratory experiment is certainly difficult, our experiment aims at comparing whether carrots or sticks lead to more cooperation in longer sequences of games than normally studied. In addition, we investigate whether carrots lead to more group solidarity by letting subjects play an additional one-shot "person-to-group" Dictator Game (DG), after the repeated PGGs. This one-shot DG – explained in details in the experimental design section – constitutes our individual measurement of group solidarity. Therefore, assuming that actors' attachment and solidarity to the group is affected by the sanctions that they receive, we postulate the following hypotheses:

H4: Individuals show lower solidarity towards other group members, the more they have been punished.

H5: Individual show higher solidarity towards other group members, the more they have been rewarded.

18.3 The experimental design

To test the hypotheses a series of computerized experiments was conducted with z-Tree (Fischbacher 2007) in the Experimental Laboratory for Sociology and Economics of a large Dutch University. Eight sessions took place at the end of 2009 with a total of 152 subjects. The subjects were students from various faculties, recruited via the online recruiting system ORSEE (Greiner 2004).

Like in Sefton, Schupp, and Walker (2007), the PGG was played as a repeated game. This characteristic allows for reciprocity in the sense that subjects can react to their fellow group members' investments either directly by sanctioning (punishing or rewarding, depending on the treatment), or indirectly, by adjusting their own investments accordingly in the next round. Furthermore, to simulate cooperation problems in groups as realistically as possible, the subjects did not know in advance when the game would end, since the end of a repeated interaction is generally unknown in real

life situations. The interactions were anonymous and, unlike in Rand et al. (2009), upon receiving a sanction subjects could not identify who had sanctioned them, so that mutual rewarding (or punishing) between rounds was impossible.

Each of the eight experimental sessions lasted between 45 and 75 minutes, depending on the treatment. Each session consisted of two series of repeated PGGs (thus, two super-games), played in the same experimental treatment. At the beginning of each session, subjects played two trial rounds to get acquainted with the procedures of the game. In these trial rounds, which had no influence on the subjects' earnings, the behaviors of the others in each group were pre-programmed. In the first trial round, the subject was informed that the other (virtual) group members had invested four points each, in the second trial round six points each. The pre-programming took place to guarantee that all subjects started the experimental rounds with the same experience. The lowest and highest possible investments, such as zero or ten points, were avoided to prevent any suggestions for extreme actions. When all individuals had completed the trial rounds, they were randomly matched into groups of four and the second part of the experiment began. The first of two super-games was played, which consisted of between 20 and 30 rounds of PGGs. After round 21 the game stopped with a probability of 10%, after round 22 with a probability of 20%, and so on until the game would stop with certainty in round 30.

Each round consisted of two parts, a PGG stage, followed by a sanctioning stage, except for the control treatment that had no sanctioning stage. At the beginning of the PGG stage, the subjects received an initial endowment of ten points and they could decide how much of this endowment to invest into a group account. Whatever the subjects did not invest into the group account was stored directly into the subject's individual account to be paid eventually at the end of the experiment. The amount collected in the group account was multiplied by three and then equally distributed between the four group members. In the sanctioning stage, the subjects received five additional points, which they could use, at a cost to effect ratio of 1:2, to punish and/or reward the other group members. That is, if a subject spent, for example, one point of his/her five points on punishing/rewarding someone else, two points were subtracted from/added to the target's payoff. All points that were not spent in sanctions were added to the subject's individual account. This manipulation of the sanctioning option is similar to the one adopted by Sefton, Schupp, and Walker (2007). However, we made sanctions more cost-effective, by adding/subtracting two points, instead of one, per every point spent on reward/punishment. This modification was implemented in order to better test our hypothesis concerning the advantage of carrots over sticks (Hypothesis 2a), because the main argument supporting that hypothesis is that rewards can increment earnings. However, as stated above, in Sefton, Schupp, and Walker (2007) rewards were simply a zero-sum transfer of points between players and, therefore, they were not economically attractive.

The experiment had four treatments and we conducted two sessions per treatment. The first treatment was the Baseline Treatment, which had no options to sanction.

The second was the Sticks Treatment, in which, in the sanctioning stage, subjects could spend any of their five points available for sanctions on punishments only. In the third treatment (Carrots Treatment) all group members could spend any of their five points available for sanctions on rewards only. In the fourth treatment both punishment and rewards were possible (Sticks & Carrots Treatment). However, the endowment at disposal remained at five points in total for both kinds of sanctions. Each subject could not spend more than the five points provided for punishing and/or rewarding. The individuals could not use their own earnings to sanction others because otherwise, subjects who earned more in the PGG might have been more willing to spend points on punishing and/or rewarding others than subjects who earned little. Thus, the measure of how willing people are to punish defectors and reward cooperators would have been biased by the amount earned in the PGG. After finishing the first super-game, all subjects were randomly re-matched in different four-person groups and the second super-game started. Throughout each super-game, the subjects were informed about the individual investments of the other members of their group by a table located at the lower part of the computer screen. Thus, subjects could make their sanctioning decisions contingent on the contributions of their fellows groups members. Furthermore, the subjects were informed about the amount that they had invested to inflict sanctions on others in each round, and about the losses and gains that were inflicted on them by punishments and rewards received from the other group members (if applicable).

Twenty subjects took part in each session, except for two sessions which had only 16 subjects (one in the Baseline and one the Carrots Treatment). Therefore, five (or four) groups of four persons were formed in each session. Each subject was seated in front of a computer station, which was visually separated from the others. As a start, the instructions about the procedures of the experiment were handed out, the subjects could choose between English and Dutch.[4] Of the 152 subjects 104 (68.4%) chose the Dutch instructions and 48 (31.6%) the English version.

Each of the two super-games per session was directly followed by a person-to-group DG which serves as our behavioral measure of a subject's solidarity towards the group. The person-to-group DG works as follows: each subject receives an endowment of 40 points. Then, he or she can decide to give some of these 40 points to the group. The amount chosen is equally distributed between all three remaining group members, without being multiplied by any number beforehand. The subjects may choose to give zero points to their fellow group members, in which case the 40 points are added to their own earnings. We used the amount given as an individual measure of solidarity towards one's own group. Finally, the experiment ended with a questionnaire asking for demographic characteristics (such as gender, age, nationality, field of studies, amount of money at disposal each month).

4 Instructions can be obtained contacting the authors.

The earnings of each subject were determined by the contributions accumulated in the group account – in every round of the PGG – divided by four (all group members) plus the rest of the endowments (from the PGG and the sanctioning part of each round) that subjects did not invest in the PGG or in sanctions, plus the points received as rewards and minus the points lost due to punishment. In addition, the points kept as well as the points received by fellow group members in the DG were likewise added to the individual account. At the end of the experiment, the subjects were paid in Euros at an exchange rate of 125 points = 1 Euro.

Of the 152 subjects who participated, 36 played in the Baseline Treatment, 40 in the Sticks Treatment, 36 in the Carrots Treatment, and 40 in the Sticks & Carrots Treatment. In total, 7568 decisions were recorded; 304 of these in the trial rounds. Thus, we recorded 7264 decisions that had influence on the subjects' earnings. Between 21 and 27, on average 23.9 rounds, were played in all super-games. The subjects earned between 5.5 and 16.5 Euros, the lowest earnings were received in the Baseline Treatment. The average profit was 12.60 Euros.[5]

18.4 Methods

For the hypotheses on cooperation, the average of the *contributions* per group serves as the dependent variable. The four different treatments serve as independent variables. For the analysis, we estimated a panel regression to account for the repeated observations within subject over time. We only used the first 20 rounds of the super-games to discard possible endgame effects. We did alternative analyses such as (1) including the final rounds, (2) using an interval regression model to take into account that groups could not contribute less than 0 or more than 40, and (3) only analyzing the first super-game. In all these analyses the pattern of results shown below is robust.

The dependent variable *solidarity* is measured by the individual contributions to the group in the *DGs*, following every super-game. The four different treatments serve as independent variables. The other independent variables are: the extent to which individuals contributed to the public good on average, the individual average received punishments, and individual average received rewards. To test the hypotheses on group solidarity we estimated a non-hierarchical mixed effects model, which includes random effects for groups and subjects. This mixed effects model is cross-classified because most subjects are involved in two different groups. The model allows us to control for possible effects of individual differences and group differences.

5 Due to a programming error, rematching was unsuccessful in two sessions. Therefore, the second super-game in these two sessions had to be excluded from the analyses presented below.

18.5 Results

18.5.1 Descriptive results

Over all treaments the contribution level was on average quite high: 8.2 points out of ten possible. Looking at the contribution levels in the different treaments, it is clear that cooperation was higher in the treatments in which sanctions were possible compared to the Baseline Treatment. The Carrots Treatment had the lowest average contribution level among the treatments with sanctions. In the two treatments with punishment, subjects assigned on average 0.35 and 0.38 punishment points to the three other subjects. Actually, in 87% of the rounds in which they could punish, subjects did not punish at all. Rewards were more frequently given, namely in 66% of the rounds in which subjects could give rewards. The average amount of points that subjects used for sanctions was 2.37 points in the Carrots Treatment and 2.03 points in the Sticks & Carrots Treatment. In the DG, the amount given to the other group members was on average 11.6 of 40 possible points, which is relatively low, compared to the percentages invested in the PGG. Table 18.1 summarizes the descriptive statistics.

Table 18.1: Descriptive statistics.

	Range	Mean	S.D.	Number of obs.
Number of rounds played	12–27	24.11	1.75	68
Contribution level in groups (overall)	0–40	36.74	6.43	1640
Individual contributions (overall)	0–10	9.18	2.16	6560
Contribution in Baseline	0–10	8.15	3.23	1384
Contribution in Sticks	0–10	9.69	1.26	1920
Contribution in Carrots	0–10	8.80	2.36	1336
Contribution in Sticks & Carrots	0–10	9.69	1.23	1920
Punishment given in Sticks	0–5	0.38	1.18	1920
Punishment given in Sticks & Carrots	0–5	0.35	0.97	1920
Punishment received in Sticks	0–12	0.38	1.09	1920
Punishment received in Sticks & Carrots	0–14	0.35	1.04	1920
Rewards given in Carrots	0–5	2.37	1.82	1336
Rewards given in Sticks & Carrots	0–5	2.03	1.89	1920
Rewards received in Carrots	0–8	2.37	1.56	1336
Rewards received in Sticks & Carrots	0–9	2.03	1.62	1920
Contributions in Dictator Game	0–40	11.63	13.74	268
Age	17–40	21.44	3.00	152
Earnings	5.5–16.5	12.60	2.72	152

As an overview, the contribution levels in each treatment are shown in Figure 18.1. This figure shows the average contribution level in each round of the 68 groups of four subjects who participated in the study. Figure 18.1 makes evident that contributions are the highest in the treatments that include sticks. Overall, the contributions in this experiment are considerably larger than in other public good experiments. This might be due to the large number of rounds subjects played. As mentioned in the description of the experiment, the super-games ended at random between 20 and 30 rounds. Therefore, the dots corresponding to rounds after the 20th represent fewer observations. For descriptive purposes we still show these rounds. They show that there are no endgame effects at all for the treatments that include sticks, while the contributions seem to decline after round 20 in the other two treatments. Furthermore, Figure 18.1 shows that the subjects in all treatments started with relatively high contributions compared to earlier experiments and even rose thereafter. However, only in the Sticks and Sticks & Carrots Treatments very high contribution levels, up to full cooperation, were maintained until the end. In the Carrots Treatment the cooperation levels were lower, and in the Baseline Treatment the contributions fell, after an initial rise, even below the starting level of cooperation.

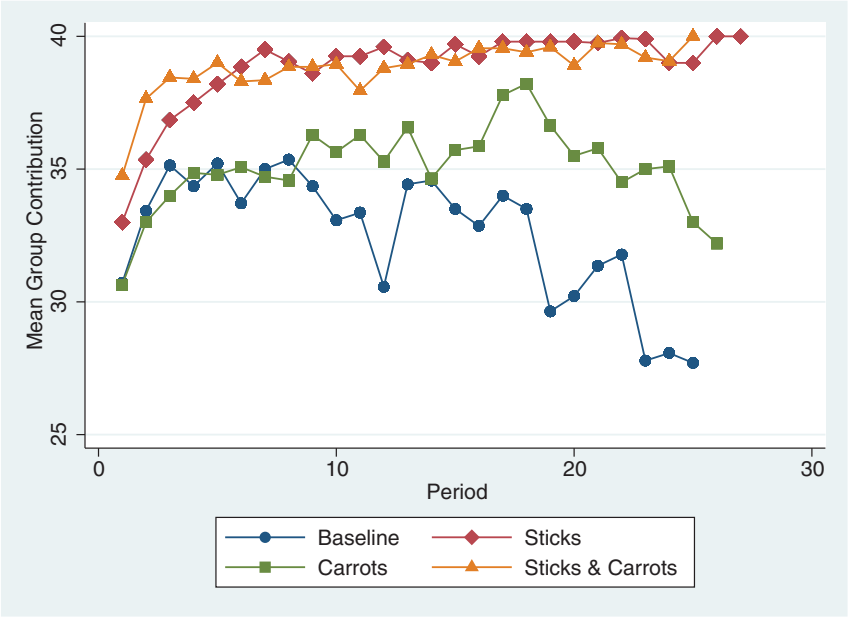

Figure 18.1: Contribution levels in the groups by treatment.

If we analyze the two super-games separately, we find that in the second super-game the subjects start in all treatments with higher contributions than in the first super-game.

Moreover, in the second super-game almost perfect cooperation is reached quicker in the Sticks and Sticks & Carrots Treatments and is stable (from round 5 onwards). The contribution levels in the Carrots and the Baseline Treatments are likewise higher than in the first super-game, but still lower than in the treatments where punishment is possible. Summarizing, the few changes that we implemented compared to the setup used by Sefton, Schupp, and Walker (2007), namely increasing the cost to effect ratio of sanctions from 1:1 to 1:2, letting subjects play more rounds, and making the end of the game probabilistic, apparently made the reputation argument in repeated games more salient, driving up contribution levels.

18.5.2 Explanatory results

18.5.2.1 Contribution to the public good

The panel regression on cooperation levels shows that contributions were significantly higher in the two treatments in which punishment was allowed (*sticks* and *sticks & carrots*) compared to the *baseline* where no sanction was allowed (see Table 18.2, Model 1). This effect is not significant for the Carrots Treatment, although the coefficient is

Table 18.2: Two-level regression on contribution with clustering on groups.

	Model 1		Model 2	
	Coefficient	**s.e.**	**Coefficient†**	**s.e.**
Baseline (reference)				
Sticks	5.21**	1.54	5.04**	1.54
Carrots	1.95	1.67	1.79	1.67
Sticks & Carrots	5.28**	1.54	5.16**	1.63
Round (Baseline)			–0.11**	0.04
Round (Sticks)			0.22**	0.03
Round (Carrots)			0.21**	0.04
Round (Carrots & Sticks)			0.12**	0.03
Constant	33.35**	1.18	34.53**	1.25
Random part				
Group level (st. dev.)	4.34		4.35	
Obs. level (st. dev.)	3.83		3.69	
R^2 within groups	0.00		0.07	
R^2 between groups	0.21		0.21	
R^2 overall	0.13		0.16	
N (groups)	*68*		*68*	
N (observations)	*1360*		*1360*	

Notes: * $p<0.05$, ** $p<0.01$ (two-sided)
† Main effects of conditions calculated for round 10

positive. Thus, Hypothesis 1 is only partly supported: cooperation levels are higher if negative sanctions can be used, but not if only positive sanctions can be used. If we include the time trends for each treatment (Model 2), we see that the differences between treatments are mainly due to increased contributions in the treatments with sanction opportunities and most strongly in treatments with sticks, while contributions decreased in the Baseline Treatment. Including other control variables, such as age and gender, does not change much in the model, but it only slightly reduces the effect sizes of the dummies for *sticks* and *sticks & carrots* and of the round variables. Thus, the model including the controls is not shown here.

Concerning Hypotheses 2a and 2b, about the comparison of contribution levels between *sticks* and *carrots*, the former is rejected and the latter is supported. That is, contribution levels are significantly higher when only punishments can be used, compared to when only rewards can be used. Finally, there are no significant differences in the contribution level between the Sticks Treatment and the Sticks & Carrots Treatment. However, the difference between the Carrots Treatment and the Sticks & Carrots Treatment is significant ($p = 0.0310$). Thus, Hypothesis 3 is only partly confirmed: if punishments and rewards are allowed simultaneously the contribution level is significantly higher than when only rewards are allowed, but it is not different from the contribution achieved when only punishment is available. Thus, the final message is clearly that sticks are more effective than carrots at sustaining cooperation and when sticks are available, the contribution level does not change, whether carrots are also available as well or not.

18.5.2.2 Solidarity

The amount given in the person-to-group DGs is used to operationalize solidarity. In Table 18.3, Model 1 shows no significant differences between any of the treatments. In Model 2, though, we can see that subjects gave significantly less to the group in the DG if they played the previous super-game in the Carrots Treatment *and* they did not receive any rewards. However, if individuals did receive rewards, they significantly gave more to the other group members in the DG. These two effects cancel out each other, because there is no main effect of carrots on solidarity. Apparently, subjects expect to receive rewards if rewards are possible. If they do not obtain carrots when they could, their solidarity towards the group diminishes. To compensate for this, subjects should receive 2.12 (= −7.47/3.52) points of rewarding per round. From Table 18.1 we know that subjects indeed on average gave a little more than 2 units in rewards per round. This explains why the main effect is indeed about zero. We did some additional analyses to see whether the experiences in the latest rounds might have had a relatively large effect on the behavior in the DG. Therefore, we distinguished between the experiences in the first 20 rounds with

Table 18.3: Mixed effects regression on contribution in Dictator Game.

	Model 1		Model 2	
	Coefficient	**s.e.**	**Coefficient**	**s.e.**
Baseline (reference)				
Sticks	3.08	2.98	1.70	3.09
Carrots	1.53	3.13	−7.47*	3.55
Sticks & Carrots	2.67	2.98	−5.91	3.43
Average contribution			0.18	0.16
Average received punishments			0.47	1.45
Average received rewards			3.52**	0.81
Constant	9.51**	2.21	3.63	5.46
Random effects				
Group level (st. dev.)	1.70		0.00	
Subject level (st. dev.)	10.94		10.53	
Obs. Level	8.03		7.92	
Log restricted-likelihood	−1041.52		−1029.43	
N (groups)	*68*		*68*	
N (individuals)	*152*		*152*	
N (observations)	*268*		*268*	

Note: *p<0.05, **p<0.01 (two-sided)

the experiences in the last couple of rounds. There were no significant differences between these effects.

Therefore, we conclude that Hypothesis 4 cannot be confirmed, since the results on the Sticks Treatments are not significant. Punishments do not lead to a negative attitude towards the group and there is not less solidarity in the punishments treatments. One explanation for this is that punishments were hardly used and, in line with the "multiplication effect" argument (Dari Mattiacci and De Geest 2009), cooperation was very high in both treatments with punishment. Thus, punishment worked as a credible threat. Conversely, Hypothesis 5 is partly supported: the individual solidarity towards the group is heightened by actual received rewards. It appears that the sheer possibility of receiving rewards leads to the expectation that one *should* receive rewards. If this expectation is not fulfilled, cooperation even decreases in the treatment with rewards.

18.6 Discussion and conclusion

The aim of this chapter was to investigate whether positive or negative sanctions in public good type of situations are more effective at promoting cooperation. Moreover,

We looked at the consequences that positive and negative sanctions produce in terms of group solidarity. We addressed these two research problems by conducting a computerized experiment. Our main findings show that cooperation levels are generally higher if sticks or sticks and carrots are allowed (Hypothesis 1). We also found that average contribution levels were rather high in our study. Moreover, contribution levels were higher in the condition in which only sticks were allowed than in the condition in which only carrots were allowed (Hypothesis 2b). When both sanctions were allowed simultaneously (Hypothesis 3), the contribution levels were significantly higher than in the condition with rewards only, but similar to the condition with punishments only.

The results on group solidarity do not lend support to the hypothesis that individuals care less about their group after they are punished (Hypothesis 4). However, if the subjects receive rewards they indeed display a higher group solidarity, as they give more to the group in the DG (Hypothesis 5). By contrast, the mere possibility to use sanctions, whether positive or negative, does not have an effect on group solidarity. Surprisingly, receiving punishment does not affect group solidarity. However, this result might be due to the low amount of actual punishment observed in our experiment. Thus, our results do not allow a strong test of the potentially demotivating effects of implemented punishments. Conversely, rewards were used more often and, as expected, produced the effect of increasing group solidarity. In addition, subjects who did not receive rewards in the condition where rewards were possible exhibited lower solidarity to the group than subjects in other conditions. Consequently, our data do not support the argument that carrots can sustain cooperation in social dilemmas (Rand et al. 2009), nor the argument that sticks have detrimental side effects in terms of group solidarity (cf. Fehr and Rockenbach 2003; Mulder et al. 2006). However, the finding that receiving rewards increases group solidarity gives some indication that cultural norms based on rewards may benefit the group in the long term, if we assume that higher group solidarity makes groups more resilient in the context of intergroup competition (cf. Richerson, Boyd, and Henrich 2003).

Our experimental setup was designed to simultaneously address some limitations of the designs used by Rand et al. (2009) and Sefton, Schupp, and Walker (2007). Our results are consistent with the idea that the high payoffs in a PGG with rewards found in Rand et al. (2009) could be due to the mutual exchange of highly efficient rewards, rather than being a pure effect of rewards on cooperation (Milinski and Rockenback 2012). By contrast, as we made the (monetary) consequences of sanctions bigger than in Sefton, Schupp, and Walker (2007), rewards were used more often than punishment by our subjects while punishment was used more often in Sefton, Schupp, and Walker (2007). Yet, we found higher contribution levels in the punishments only treatment than in the rewards only treatment, suggesting that sticks function as credible threats, i.e., they support cooperation without needing to be used (cf. Dari Mattiacci and De Geest 2009). Our experimental design also implemented a higher number of within group interactions than did Sefton, Schupp, and

Walker (2007), coupled with an uncertain end of the super-game. As a result, we obtained relatively high levels of cooperation, even in the treatment without sanctions. Consequently, our subjects provided more room for the use of positive than negative reinforcers. Due to the 1:2 cost to effect ratio, in our rewards condition, Pareto-superior outcomes could be reached if everyone gave rewards to everyone all the time. Perhaps some of our subjects realized this and others did not. If the actors who did realize this expected that indeed everyone exchanged rewards, they may have been disappointed when others did not conform to this expectation, because they lost points from providing rewards that they did not receive back from others. This last point is clearly related to the costs of rewards. The actors who provided rewards would not have been so much bothered if the rewards were for free. This implies that the cost argument of Dari-Mattiacci and De Geest (2009) is relevant in this context. The threat of sticks is effective and cheap, while the costs of continuously providing carrots make them less effective.

Our results also imply that more theoretical work about group solidarity is needed. For example, the negative effect of sticks, e.g., reduced group solidarity, might not be generated by the possibilities of punishments alone, but it might require a context in which punishment is necessarily applied. This could be modeled as a public good with uncertainty, i.e., a setting in which the contribution to the public good is not always perfectly visible and sometimes actors do not seem to contribute, while they actually do. Finally, the weak results concerning the effects of sanctions on solidarity might also be due to our measurement of group solidarity as a one-shot person-to-group DG. The behavior in the DG can be determined by many subtle cues in the PGGs. In future studies, it would be advisable to complement the findings with more detailed measurements of solidarity and attachments to the group, for example, using some attitudinal scales, next to behavioral measures such as the DG.

References

Andreoni, James. 1988. "Why Free-Ride?: Strategies and Learning in Public Goods Experiments." *Journal of Public Economics* 37: 291–304.

Andreoni, James, William Harbaugh, and Lise Vesterlund. 2003. "The Carrot or the Stick: Rewards, Punishments, and Cooperation." *American Economic Review* 93: 893–902.

Axelrod, Robert. 1984. "*The Evolution of Cooperation*." New York: Basic Books.

Balliet, Daniel, Laetitia Mulder, and Paul A. M. van Lange. 2011. "Reward, Punishment, and Cooperation: A Meta-Analysis." *Psychological Bulletin* 137: 594–615.

Barclay, Pat. 2006. "Reputational Benefits for Altruistic Punishment." *Evolution and Human Behaviour* 27: 325–344.

Boyd, Robert J., Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Science* 100: 3531–3535

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. "Bad Is Stronger Than Good." *Review of General Psychology* 5: 323–370.

Bohnet, Iris, and Yael Baytelman. 2007. "Institutions and Trust: Implications for Preferences, Beliefs and Behavior." *Rationality and Society* 19: 99–135.

Buskens, Vincent, and Werner Raub. 2013. "Rational Choice Research on Social Dilemmas: Embeddedness Effects on Trust." In *Handbook of Rational Choice Social Research*, edited by Rafael Wittek, Tom A. B. Snijders, and Victor Nee, 113–150. Stanford: Stanford University Press.

Calvert, Randall L. 1995. "Rational Actors, Equilibrium, and Social Institutions." In *Explaining Social Institutions*, edited by Jack Knight and Itai Sened, 57–94. Ann Arbor, MI: University of Michigan Press.

Chauduri, Ananish. 2011. "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature." *Experimental Economics* 14: 47–83.

Dari-Mattiacci, Giuseppe, and Gerrit De Geest. 2009. "Carrots, Sticks, and the Multiplication Effect." *Journal of Law, Economics and Organization* 73: 377–386.

Fehr, Ernst, and Urs Fischbacher. 2003. "The Nature of Human Altruism." *Nature* 425: 785–791.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415: 137–140.

Fehr, Ernst, and Herbert Gintis. 2007. Human Motivation and Social Cooperation: Analytical and Experimental Foundations. *Annual Review of Sociology* 33: 43–64.

Fehr, Ernst and Joseph Henrich. 2003. "Is strong Reciprocity a Maladaptation? On the Evolutionary Foundations of Human Altruism." In *Genetic and Cultural Evolution of Cooperation*, edited by Peter Hammerstein, 55–82. Cambridge, MA: The MIT Press.

Fehr, Ernst, and Bettina Rockenbach. 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature* 422: 137–140.

Fehr, Ernst, and Klaus, M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economic* 114: 817–68.

Fischbacher, Urs. 2007. "z-Tree – Zurich Toolbox for Readymade Economic Experiments – Experimenter's Manual." *Experimental Economics* 10: 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economics Letters* 71: 397–404.

Flache, Andreas. 2002. "The Rational Weakness of Strong Ties. Failure of Group Solidarity in a Highly Cohesive Group of Rational Agents." *Journal of Mathematical Sociology* 26: 189–216.

Friedkin, Noah E. 2004. "Social Cohesion." *Annual Review of Sociology* 30: 409–25.

Gächter, Simon. 2013. "Rationality, Social Preferences and Strategic Decision-Making from a Behavioral Economics Perspective." In *Handbook of Rational Choice Social Research*, edited by Rafael Wittek, Tom A. B. Snijders, and Victor Nee, 33–71. Stanford: Stanford University Press.

Gintis, Herbert. 2000. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206: 169–179.

Greiner, Ben. 2004. "The Online Recruitment System ORSEE 2.0. A Guide for the Organization of Experiments in Economics." University of Cologne, Working Paper Series in Economics 10.

Hamilton, William D. 1963. "The Evolution of Altruistic Behavior." *The American Naturalist* 97: 354–356.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment across Societies." *Science* 319: 1362.

Hobbes, Thomas. 1651. *Leviathan or the Matter, Forme and Power of a Commonwealth, Ecclesiasticall and Civill*. Oxford: Clarendon Press, 1909, reprinted from the edition of 1651, London.

Irwin, Kyle, Laetitia Mulder, and Brent Simpson. 2014. "The Detrimental Effects of Sanctions on Intragroup Trust: Comparing Punishments and Rewards." *Social Psychology Quarterly* 77: 253–272.

Johnson, Dominic P., Pavel Stopka, and Stephen Knights. 2003. "The Puzzle of Human Cooperation." *Nature* 421: 911–912

Keser, Claudia, and Frans van Winden. 2000. "Conditional Cooperation and Voluntary Contributions to Public Goods." *Scandinavian Journal of Economics* 102: 23–39.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory* 27: 245–252.

Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth, 111–194. New York: Princeton University Press.

Lindenberg, Siegwart. 2001. "Social Rationality versus Rational Egoism." In *Handbook of Sociological Theory*, edited by John H. Turner, 635–668. New York: Kluwer Academic/Plenum Publishers.

Markovsky, Barry, and Edward J. Lawler. 1994. "A New Theory of Group Solidarity." *Advances in Group Processes* 11: 113–137.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review* 93: 366–380.

Milinski, Manfred, and Bettina Rockenbach. 2012. "On the Interaction of the Stick and the Carrot in Social Dilemmas." *Journal of Theoretical Biology* 299:139–143.

Mulder, Laetitia B., Eric Van Dijk, David De Cremer, and Henk A. M. Wilke. 2006. "Undermining Trust and Cooperation: The Paradox of Sanctioning Systems in Social Dilemmas. *Journal of Experimental Social Psychology* 42: 147–162.

Nikiforakis, Nikos. 2008. "Punishment and Counter-punishment in Public Goods Games: Can We Really Govern Ourselves?" *Journal of Public Economics* 92: 91–112.

Nowak, Martin A. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314: 1560–1563.

Nowak, Martin A, Anna Dreber, David G. Rand, and Drew Fudenberg. 2008. "Winners Don't Punish." *Nature* 452: 348–351.

Nowak, Martin A., and Karl Sigmund. 1998. "Evolution of Indirect Reciprocity by Image Scoring." *Nature* 393: 573–577.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 58: 1356–1375.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

Pan, Xiaofei, and Daniel Houser. 2017. "Social Approval Competition and Cooperation." *Experimental Economics* 20: 309–332.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83: 1281–1302.

Rand, David G., Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin A. Nowak. 2009. "Positive Interactions Promote Public Cooperation." *Science* 325: 1272–1275.

Raub, Werner, Vincent Buskens, and Rense Corten. 2015. "Social Dilemmas and Cooperation." In *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, edited by Norman Braun and Nicole J. Saam, 597–626. Wiesbaden: Springer VS.

Raub, Werner, and Jeroen Weesie. 1990. "Reputation and Efficiency in Social Interactions: An Example of Network Effects. *American Journal of Sociology*, 96: 626–654.

Richerson, Peter J., Robert T. Boyd, and Joseph Henrich. 2003. "Cultural Evolution of Human Cooperation." In *Genetic and Cultural Evolution of Cooperation*, edited by Peter Hammerstein, 357–388. Cambridge, MA: MIT Press.

Sefton Martin, Robert Shupp, and James M. Walker. 2007. "The Effects of Rewards and Sanctions in Provision of Public Goods." *Economic Inquiry* 45: 671–690.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87: 355–374.

Stigler, George J., and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum." *American Economic Review* 67: 76–90.

Stoop, Jan, Daan Van Soest, and Jana Vyrastekova. 2018. "Rewards and Cooperation in Social Dilemma Games." *Journal of Environmental Economics and Management* 88: 300–310.

Takacs, Karoly, and Bela Janky. 2007. "Smiling Contributions: Social Control in a Public Goods Game with Network Decline." *Physica A* 378: 76–82.

Tenbrunsel, Ann E., and David M. Messick. 1999. "Sanctioning Systems, Decision Frames, and Cooperation." *Administrative Science Quarterly* 44: 684–707.

Traulsen, Arne, and Martin A. Nowak. 2006. "Evolution of Cooperation by Multilevel Selection." *Proceedings of the National Academy of Science* 103: 10952–10955.

Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46: 35–57.

Willer, Robb. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74: 23–43.

Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51: 110–116.

Zahavi, Amos. 1995. "Altruism as a Handicap – the Limitations of Kin Selection and Reciprocity." *Journal of Avian Biology* 26: 1–3.