Exit, punishment and rewards in commons dilemmas: An experimental study

Giangiacomo Bravo,* Flaminio Squazzoni†

To be presented at the conference:
"Design and Dynamics of Institutions for Collective Action"
A Tribute to Prof. dr. Elinor Ostrom (1933-2012)

Abstract

This paper investigates the interplay of incentives, sanctions and voluntary participation in commons dilemmas. We performed a modified public good game experiment to examine the effect of exit, rewards and punishment, as well as that of the interplay between exit and rewards and punishment, on cooperation. We found that punishment had a stronger effect than rewards on cooperation if singularly considered, whereas rewards had a stronger effect when combined with voluntary participation. This can be explained in terms of "frame effect" as the combination of exit and rewards might induce people to attach higher expected payoffs to cooperative strategies and expect more good behaviour from other individuals.

KEYWORDS: social dilemmas; incentives; rewards; punishment; cooperation; exit.

1 Introduction

Commons dilemmas are interaction situations where a common good is provided or exploited by a population of individuals so that optimal collective outcomes are at odds with private interests (Hardin 1968; Ostrom 1990, 1998). Self-interest determines that individuals might behave in ways that are socially counter-productive. In all cases of public good provision, such as good quality peer review in science or a peaceful and clean public beach in a popular place, a multitude of individuals are called to put their private resources into the public pot to the benefit of the whole group, including those who did not contribute. In cases of common-pool resource (CPR), such as natural resource exploitation or artificial infrastructure use, individuals draw benefits by sharing a good where significant externalities of consumption exist.

Although any collective outcome can be maximised when everyone cooperates, the only self-interest rational strategy for each individual is to free-ride, either by contributing nothing to the public good or by extracting more than the sustainable share from the

*Department of Economics and statistics, University of Torino, and Collegio Carlo Alberto <giangiacomo.bravo@unito.it>.

†GECS Research Group, Department of Economics and Management, University of Brescia <squazzon@eco.unibs.it>.

common resource, predicting that others will do the same. This is a source of socially counter-productive outcomes. For instance, it could be individually rational to use private transportation for daily trips as this potentially reduces the time needed to get to the workplace. On the other hand, this decision rises traffic congestion and pollution, with all negative consequences on health and time wasted in traffic jams.

In the social and economic life, institutions and organizations provide material and non-material incentives to help people to overcome their social dilemmas (North 2005; Ostrom 2005; Whitmeyer 2002). Various institutional arrangements, such as bonus programmes and ethical codes in organizations, can be seen as targeted interaction that embodies a coherent set of rewards and punishments, which can make cooperation more predictable (Tenbrusel and Messick 1999; Okhuysen and Eisenhardt 2002). Similarly, commons management institutions regulate resource exploitation by posing limits to individual consumption and punishing overuse (Ostrom 1990, 2005).

In this vein, a recent study by Sasaki et al. (2012) looked at the role of incentives and sanctions and especially at the interplay of incentives provided by an institution and the effect of voluntary participation in public goods games. This study applied evolutionary game theory to respond to certain limitations of previous literature. On the one hand, previous studies on the interplay of incentives and sanctions did not achieve unequivocal findings. A well-known experimental overview (Andreoni et al. 2003) showed that there is substantial demand both for rewards and punishments in many empirical situations. Their results indicated that cooperation is less probable when good behaviour is rewarded as to where institutional arrangements exist that punish bad behaviour. By experimentally investigating the interplay and the economic efficiency of rewards and punishment in public good games, (Sefton et al. 2007) concluded that certain synergies between the two measures could even take place, although punishment is more effective in promoting cooperation and even easily institutionally build-in than rewards, but to detriment of efficiency (see also Dari-Mattiacci and Geest 2010; Milinski and Rockenbach 2012).

This contrasts with the experimental results of Tenbrusel and Messick (1999), where it was found that punishment could even have a negative effect on cooperation as it would induce individuals to frame the game as a self-interest competition situation with defection as an expected dominant strategy, especially in case of weak and poorly credible sanctions (e.g. Kahneman and Tversky 2000). By introducing the possibility that players' identity was known in repeated public good games, Rand et al. (2009) showed that this positive effect of rewards against punishment could be significantly increased.

While voluntary participation is often considered positive in organizational literature (e.g., Tenbrusel and Messick 1999; Okhuysen and Eisenhardt 2002; Spicer 1985), studies on CPR management usually consider any exit option as a factor reducing interdependence within the users' group and dependence from the resource (e.g., Agrawal 2002). This means that exit can have a negative effect on cooperation in the commons. Moreover, little is known on the interplay between exit option and incentives (Sasaki et al. 2012). On the one hand, rewards, punishment and exit options could be singularly viewed as alternative, or at best synergistic measures to motivate people and improve individual effort and commitment for cooperation (e.g., Spicer 1985). Indeed, voluntary participation may drive individual attention towards freedom and willingness and so promote self-motivated good behaviour (e.g., Tenbrusel and Messick 1999; Okhuysen and Eisenhardt 2002). On the other hand, exit can favour free-riding by allowing individuals to escape punishment

and reducing the commitment of individual towards the group's interests (Richerson and Boyd 2001).

The aim of this paper was contribute to this recent stream of research by experimentally investigating: (i) whether voluntary participation could favour cooperation in commons dilemma situations where rewards and punishment are insufficient to establish cooperation, i.e., they are not such to change the dominant strategy of defection for players, and (ii) what is the effect of the interplay of rewards and punishment with or without exit option. The rest of the paper is structured as follows. Section 2 discusses our research background, with a particular attention to the potential effect of incentives, in forms of rewards and punishment, and exit on cooperation. Section 3 illustrates our experimental design and Section 4 presents the hypotheses that we derived from the literature. Section 5 presents the results, while in the closing section (6), we discuss our findings to draw also some implication for further work.

2 Research background

The archetypal model of a social dilemma is the Prisoner's dilemma (PD) game, where two players simultaneously decide whether to cooperate or not in a situation where cooperation leads to the social optimum, while defection (or free-riding) maximizes short-term individual benefits. In the extension of this cooperation game to n players, known as Public Good game (from now on, PG), participants are endowed with a fixed sum of money and choose whether to keep it in their own private account (= defection) or to contribute to the public good (= cooperation) (e.g., Isaac et al. 1985; Ledyard 1995). The amount kept by participants increases their payoff, while what is contributed is multiplied by a factor $m \in (1/n, n)$ and then divided evenly among everyone, independently of their contribution. Since $m < n$, it is not individually beneficial to contribute to the public good, irrespectively of what other individuals do. Therefore, the game has a dominant strategy of keeping the entire endowment. However, being $m > 1/n$, in case all participants would contribute, everyone would stay better, with a social optimum given by all participants contributing their whole amount. It is worth noting that CPR games pose a similar dilemma, although in these cases the problem is not whether to contribute to a public good but how to reduce the exploitation level from a common pool (e.g., Ostrom et al. 1994; Walker et al. 1990).

Experimental results of PG and CPR games consistently rejected the game-theoretical prediction of universal defection, with cooperation usually starting at intermediate levels. On the other hand, when the game was repeated in conditions of anonymity and without communication, cooperation progressively declined over time, approaching zero after a few rounds (Ledyard 1995; Ostrom et al. 1994). To solve this problem, a variety of factors able to counterbalance defection have been investigated that included, among others, the marginal cooperation's gain (i.e., m), group size and stability, and various communication and reputation systems (Gächter and Herrmann 2009; Ostrom 2006).

At the same time, other studies investigated the effect of sanctions, including cooperators' rewards and free-riders' punishments (e.g., Fehr and Gächter 2000, 2002; Herrmann et al. 2008; Ostrom et al. 1992; Rand et al. 2009; Sefton et al. 2007; Walker et al. 2000; Walker and Halloran 2004). In this case, any incentive greater than the cost of cooperation, whether positive or negative, should ideally change the dominant strategy

of defection at the individual level. For instance, imagine a situation where two players simultaneously choose whether to pay a cost $c > 0$ to give a benefit $b > c$ to the opponent. In this case, the structure of the game is similar to a PD with a dominant strategy of defection. Nevertheless, both a fine for defectors and a reward for cooperators, if greater than c, can change the players' behaviour and lead to full cooperation. Therefore, the mere threat of the fine is sufficient to avoid free-riding, while rewards have actually to be paid when cooperation is established, so presenting a direct cost for the organization or institution involved. This led some authors to argue that the threat of punishment is more effective than the promise of rewards. This is why both democratic and non-democratic governments largely rely on "sticks" and rarely use "carrots" to foster rule compliance (Dari-Mattiacci and Geest 2010).

Previous investigation on PG games showed that individuals are willing to punish defectors, even in one shot games or when the possibility of repeated encounters between the same players was ruled out (e.g., Fehr and Gächter 2000, 2002; Gächter and Herrmann 2009; Herrmann et al. 2008; Ostrom et al. 1992; Walker and Halloran 2004). Punishment usually takes the form of a fine that subjects can impose to other group members at a cost for themselves. For instance, after receiving information about other players' behaviour, each participant can decide whether to use part of their endowment to punish other group members. In most cases, the rule is that for each monetary unit (MU) used in punishing the target is fined by three MU (e.g., Fehr and Gächter 2000, 2002; Gächter and Herrmann 2009; Herrmann et al. 2008). However, at least in the short run, the cost of the fines overcomes cooperation gains (Fehr and Gächter 2000; Herrmann et al. 2008; Sefton et al. 2007). Although some experiments indicated that a net benefit may be obtained when the interaction is repeated a sufficient number of times (Gächter et al. 2008), punishment *decreases* participants' earnings, leaving unsolved the question of whether this institutional scheme is actually profitable and robust. Moreover, recent CPR experiments showed that punishment does not positively affect participants earnings unless combined with communication (Janssen et al. 2010).

An alternative to defectors' punishment is rewarding cooperators. In this case, each participant can use part of his/her endowment to increase the earnings of other participants, often with the usual three to one ratio. Experiments showed that participants who cooperate in PG games are inclined to reward other cooperators (e.g., Milinski and Rockenbach 2012; Rand et al. 2009; Sefton et al. 2007; Walker and Halloran 2004). Moreover, when the choice is possible, participants tend to prefer rewards over sanctions (Sutter et al. 2010). Although the fact that rewards might be more effective than punishment to sustain cooperation is still debated, results indicate that individuals usually prefer not to incur sanctions for their behaviour (Dari-Mattiacci and Geest 2010; Rand et al. 2009; Sefton et al. 2007; Sutter et al. 2010).

In this respect, a few papers investigated the effect of centralized institutions that might induce participants to cooperate. The attention has been so far concentrated more on understanding the opportunity to implement these institutional solutions. Kosfeld et al. (2009) designed a public-good experiment where participants had the possibility of implementing an external cooperation-enforcing "organization" by paying a fixed cost. They found that, even if many groups succeeded in implementing this organization and consequently achieved higher payoffs, this outcome was not robust and depended both on structural factors (e.g., the return rate from the public good and the number of group

members) and on the perceived "fairness" of the organization. Similarly, in a CPR experiment, Walker et al. (2000) found that introducing the possibility of voting for a mandatory "allocation rule" substantially increased the efficiency of the outcomes. Surprisingly, they found that requiring the unanimity as a condition to select the enforcing institution was more efficient than simply relying on a majority voting rule.

Early theoretical works iterated PD games hypothesized that voluntary participation could led to an increase of cooperation (Batali and Kitcher 1995). Subsequent studies generalized this finding to PG games and showed that, by introducing an exit option, the predominance of a single strategy was less likely than a rock-paper-scissors succession of cooperators, defectors and "loners" (agents choosing not to participate in the game) (Hauert et al. 2002; Sasaki et al. 2007). Recently, Sasaki et al. (2012) examined the interplay of institutional incentives and voluntary participation in public goods games using an evolutionary game theory approach. They found that voluntary participation can foster cooperation especially in when combined with punishing, while the combined effect of voluntary participation and rewards was weaker, leading to high cooperation levels only when incentives were considerably higher.

Experimental findings confirmed that the possibility of exit increases cooperation in one-shot PDs (Orbell and Dawes 1993). Indeed, with exit, intending cooperators were more willing to enter the game than intending defectors, thereby increasing the probability of win-win equilibria. Also in iterated PG game experiments, results showed that cooperation tends to be higher in cases of voluntary participation Semmann et al. (2003). Moreover, the oscillating dynamics between strategies predicted by theoretical models was experimentally confirmed.

3 Methods

A total of 144 subjects (58% females) participated in the experiment, which was held at the University of Brescia on April 23, 2012. Participants were students of the Faculty of Economics recruited using the on-line system ORSEE (Greiner 2004). They played in sessions of 24 subjects and interacted anonymously through a computer network using the experimental software z-Tree (Fischbacher 2007). Each experimental session took approximatively 40 minutes, including instruction reading. The average payoff, including the show-up fee, was 10.52 Euro and all earnings were paid in cash immediately at the end of the experiment.

All participants played 10 periods of an introductory "modified public good game", presented below, plus 10 further treatment periods different for each session.[1] The goal of the introductory periods was to let participants familiarize with the game and create a situation where defection dominated, while treatments allowed us to test the effect of the incentive schemes and exit.

The game was played in groups of six subjects, which changed after each period. At the beginning of the game, participants received an endowment of 100 Monetary Units (MU), with an exchange rate of 1 MU = 2 Euro cents. In each period, they were asked to decide whether to bear a cost of 10 MU to provide a benefit of 20 MU to the other group members, evenly divided among them. This means that the individual payoff in

[1] Complete instructions to participants are provided in the Appendix.

case of full cooperation was 10 MU, while that for full defection was 0 MU. However, an unilateral defector could earn 20 MU while a cooperator in an otherwise defecting group could lose 10 MU. At the end of each period, the resulting payoffs were added/subtracted to/from the endowment and the outcome was communicated to all players.

In designing the game, we followed Sasaki et al. (2012) in assuming that the contribution of each subject provide benefits only to other group members. This means that contributing in our game was a purely altruistic action, with nothing returned to the cooperative player. This makes the decision unequivocally cooperative and better approximating real-world situations—as the example of limiting private transport use presented in the introduction, where the direct benefit of cooperation is negligible due to the large number of individuals playing the dilemma.

At the end of the introductory periods, all subjects received a new endowment of 100 MU and played 10 further periods where one of the variables below was manipulated following a 2×3 factorial design. The first factor, called *Exit*, concerned an exit option allowing subjects to decide in each period whether to participate or not in the game, while in the *No Exit* treatments, all subjects played the game as before. Consistently with previous experiments that introduced the same variable, either exiting or participating in the game had no cost (Orbell and Dawes 1993; Semmann et al. 2003). Subjects who chose to participate played as before, while those who opted out bore no cost but could not derive any benefits from cooperation. Note that, since opting out reduced the number of active group members, each of the remaining players earned a higher share of the 20 MU from cooperative choices. On the other hand, in order to rule out any strategic behaviour from knowing the number of other subjects who participated in the game, each decision whether to participate and cooperate were taken simultaneously by everyone. To sum up, subjects in the exit treatments faced a three-option choice between (i) providing the benefit, (ii) not providing the benefit, and (iii) not participating in the game. Subjects were also told that, in case of only one subject choosing to participate, the current period game would not have been carried out.

The second factor, called *Incentive*, was based on three incentive levels: null, positive, and negative. Under the positive scheme, we assumed that a reward of 5 MU was awarded to cooperators. Under the negative scheme, a punishment of 5 MU was dispensed to free-riders. Under the null scheme, all subjects played the game as before. Note that the level of the incentive was intentionally such to ensure that players' dominant strategy was still to defect. Here, we assumed that rewards and punishment were established by an external controller and that perfect enforcement existed. Table 1 summarizes the six treatments depending on the intersection of all factor levels.

		Incentive		
		Null	Negative	Positive
Exit	No	*No-Null*	*No-Neg*	*No-Pos*
	Yes	*Ex-Null*	*Ex-Neg*	*Ex-Pos*

Table 1: Overview of the experimental design with treatment labels.

It is worth noting that, while experimental research mostly examined decentralized punishment (e.g., Fehr and Gächter 2002, 2000; Ostrom et al. 1992), in many real-world

situations an institution may exist, at least partially separated from individuals or organizations "that play the game", tha has the function of administering sanctions to players. In cases of private business, public administration and common-pool resources, the puzzle to be explained is not who should enforce the rule but whether the enforcement level is effective in providing sufficient incentives to overcome the free-riding temptation of individuals.

4 Hypotheses

Following the experimental results presented above, we formulated six hypotheses about the *expected* outcome of the different treatments (plus the introductory periods).

Hypotheses 1 *In the introductory periods, cooperation is expected to start at intermediate levels to subsequently decline.*

In standard public-good games, cooperation usually starts at intermediate levels and subsequently declines (Andreoni and Croson 2008; Isaac et al. 1985; Ledyard 1995). There was no reason to expect that our game would have made a difference. Being the cost of cooperation consistent, i.e., nothing of the contribution was returned to the contributor, we expected that the decline could be even more pronounced in our experiment than in standard PG games.

Hypotheses 2 *The* No-Null *treatment should lead to cooperation levels similar or lower than the introductory periods. We expect that its dynamics follows a downward trend.*

Restarting a public-good game usually leads to an increase in cooperation (even if not necessarily up to the initial level) followed by a new decline (Andreoni and Croson 2008). In *No-Null*, we expected that cooperation should decline over time like in the introductory periods. Given that participants had already experienced the game during the introductory periods, we expected that the decline in cooperation could be even more pronounced.

Hypotheses 3 *Cooperation in the* Ex-Null *treatment is expected to be higher than what observed in* No-Null.

Theoretical works (Batali and Kitcher 1995; Hauert et al. 2002; Sasaki et al. 2007) and previous experiments (Orbell and Dawes 1993; Semmann et al. 2003) showed that voluntary participation may increase cooperation. Coherently, we expected that *Ex-Null* should determine a higher proportion of cooperative move than *No-Null*, where participation was mandatory. Following Orbell and Dawes (1993), who found that, when participation is voluntary, intending cooperators are more willing to enter the game than intending defectors, we expected to observe higher exit choices by intending defectors.

Hypotheses 4 *Punishment and rewards with no exit (*No-Neg *and* No-Pos *treatments) should increase cooperation but not sufficiently to stop its downward trend.*

Although insufficient to change the dominant strategy of rational players, we expected that both punishments and rewards should increase cooperation compared with *No-Null*. This is because individuals tend to react to sanctions even when these are only symbolic (de Quervain et al. 2004) or not credible (Burnham and Hare 2007). However, following previous experimental results, we expected that the presence of a significant share of free-riders should progressively reduce cooperation (see Fehr and Gächter 2002; Gintis et al. 2003).

Hypotheses 5 *When exit is combined with punishment (*Ex-Neg *treatment), cooperation should increase more than in* No-Null *and* No-Neg *treatments.*

Under a negative incentive scheme, free-riders have a rational interest in leaving the game as long as the fees are higher than the expected benefit from cooperation of other individuals, i.e., $bn/(m-1)$ MU, where b was the benefit given to others group members, m was the number of subjects participating in the game and n was the number of cooperative subjects. Given the parameters of the game, such an outcome was expected if there were at least two cooperators in the group. The exit of free-riders is expected to allow cooperation to spread at least up to the point where free-riders have interest to return something in the game. Moreover, Sasaki et al. (2012) argued that this scheme can lead to full cooperation even with low incentives. We expected to find from intermediate to high levels of cooperation and a significant use of the exit option, manly by intending defectors.

Hypotheses 6 *When exit is combined with rewards (*Ex-Pos *treatment), cooperation should increase, but not more than in* No-Pos*.*

Under a positive incentive scheme, free-riders have no interest in leaving the game and so the exit option is non-influent. Moreover, Sasaki et al. (2012) predicted that this combination should be less effective than the voluntary participation plus punishment. Therefore, we expected cooperation levels and game dynamics similar to *No-Pos* treatment.

5 Results

In line with past PG game results, and consistently with our first hypothesis, cooperation in the introductory periods started at intermediate levels to decline subsequently, leading to a situation where defection was the most common strategy. In subsequent periods, cooperation varied depending on the treatment (Fig. 1).

Even if none of the treatments was capable of fully stopping the decline in cooperation that is typical of PG and social dilemma games, our experimental conditions led to significantly different outcomes (Fig. 2). Consistently with our second hypothesis, defection prevailed in *No-Null*, which was a repetition of the introductory periods and was our control condition. The average proportion of cooperative moves was 0.15 ± 0.02 with a declining trend leading close to zero cooperation in the final periods.

The sole introduction of voluntary participation was not sufficient to change the situation. Unlike our third hypothesis, *Ex-Null* led only to a minimal increase of cooperation. The average proportion of cooperative moves was 0.21 ± 0.03, which was not significantly different from *No-Null* (Wilcoxon rank sum test on individual averages: $W = 244.5$,

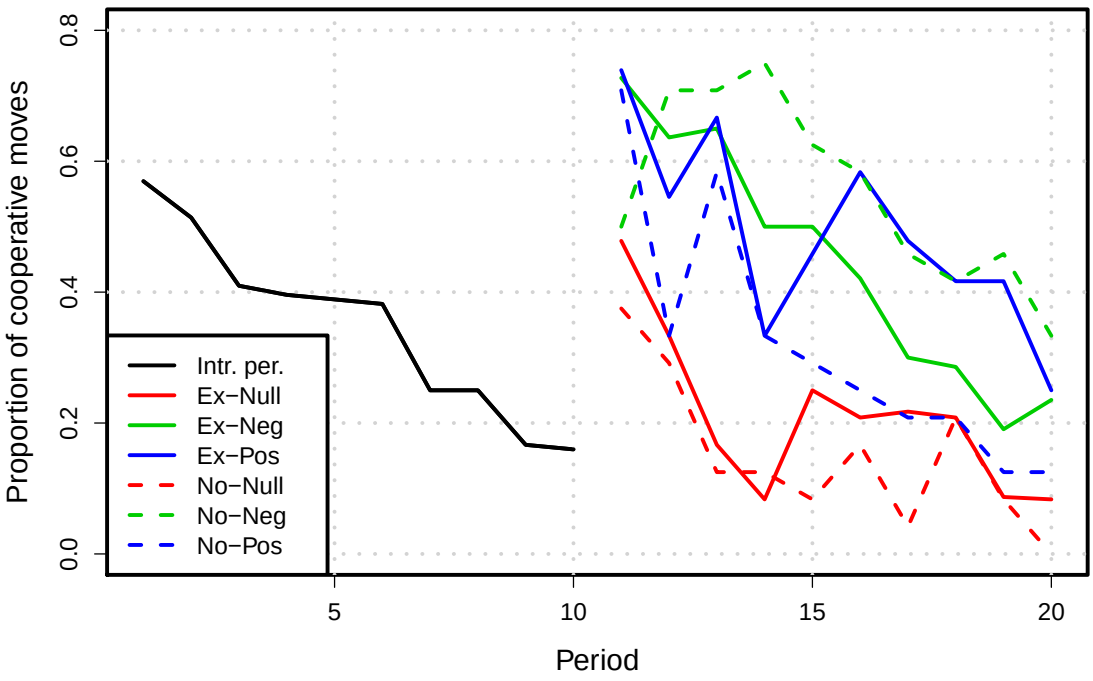

Figure 1: Average cooperation proportion per treatment and period. For the sake of readership, introductory period data for all groups were pooled in a single curve.

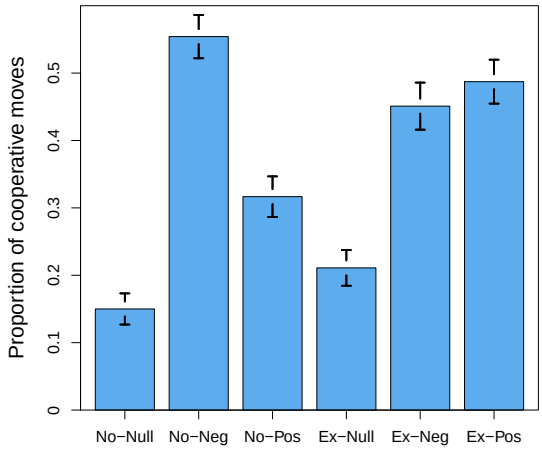

Figure 2: Average cooperation proportion per treatment with standard error bars.

$p = 0.180$ one tailed). It is also worth noting that participants rarely opted to exit, i.e., only slightly more than 1% of the times.

The outcome changed when institutional incentives were introduced. Although rewards were theoretically insufficient to alter the players' dominant strategy, *No-Pos* led

to significantly higher cooperation (0.32 ± 0.03) than *No-Null* ($W = 148$, $p = 0.002$ one tailed), even if defection still dominated, especially in the final periods. On the other hand, *No-Neg* led to a majority of cooperative moves (0.55 ± 0.03). The difference with *No-Null* was highly significant ($W = 86.5$, $p < 0.001$ one tailed) and the treatment led to significantly higher cooperation compared with *No-Pos* ($W = 157.5$, $p = 0.004$ one tailed). Therefore, in the case of mandatory participation, the influence of punishment on cooperation was higher than rewards. This is consistent with our fourth hypothesis, even if the superiority of punishments over rewards was not expected.

The introduction of voluntary participation combined with the incentive schemes generally led to more cooperation. Consistently with our fifth hypothesis, *Ex-Neg* led to higher cooperation than *No-Null* (0.45 ± 0.03, $W = 68.5$, $p < 0.001$ one tailed). However, cooperation levels were slightly lower than in *No-Neg*, although the difference was statistically significant only at the 10% level ($W = 344.5$, $p = 0.073$ one tailed). As expected, this was the treatment where most participants chose to exit (15%), with less cooperative participants who chose to exit more frequently as predicted. The correlation between the individual proportion of cooperative moves in the 10 introductory periods of the game and the number of exit in the treatment periods was negative ($r = -0.41$). This means that punishment induced intending defectors to seriously consider to opt out to avoid fees.

Ex-Pos led to a similar proportion of cooperative moves (0.49 ± 0.03). In this case, subjects rarely choose to exit (i.e., less than 2% of the times). The difference with *No-Null* was highly significant ($W = 46$, $p < 0.001$). It is worth noting that *Ex-Pos* led to more cooperation than *No-Pos* ($W = 161.5$, $p = 0.004$ one tailed). Moreover, unlike the case where participation was mandatory, in this case, the level of cooperation approached the case of negative incentives, i.e., *Ex-Neg*. It is worth noting that, while the fact that *Ex-Pos* led to more cooperation than *No-Null* was consistent with our sixth hypothesis, the fact that the treatment led to cooperation levels similar to *Ex-Neg* and above *No-Pos* was not expected.

To examine the interplay of incentives and exit in greater detail, we performed an analysis of variance on the proportion of cooperative moves for each subject in all treatment periods.[2] Results showed that exit was not significant in itself (so our third hypothesis did not hold) but had a significant effect when combined with rewards and a weakly significant effect when combined with punishment (consistently with our sixth and, to some extent, fifth hypotheses). On the other hand, unlike rewards, the pure effect of punishment was highly significant (Tab. 2).

As regards to participants' earnings, *Ex-Pos* led to the highest absolute final profit, followed by *No-Pos* and *No-Neg* (Fig. 3a). In order to control for the fact that extra-money was on stake in *No-Pos* and *Ex-Pos*, we also measured the profit as proportion of the theoretical optimum, i.e., as the amount earned in case of full cooperation plus the sum of all positive incentives (Fig. 3b). In *Ex-Pos*, participants achieved a profit equal to 69% of the optimum, which is the best result in all treatments. This means that the combination of voluntary participation and rewards not only ensured high cooperation

[2]Following other experimental papers (e.g., Barrera and Buskens 2009; Boero et al. 2009; Buskens et al. 2010; Corten and Buskens 2010), to double-check our results we also estimated logit models with random effects using a GLM panel estimator. This led to outcomes fully consistent with the results presented in the paper.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
exit	1	0.051	0.051	1.005	0.318
rewarding	1	0.127	0.127	2.470	0.118
punishment	1	2.422	2.422	47.287	0.000
exit × rewarding	1	0.300	0.300	5.865	0.017
exit × punishment	1	0.177	0.177	3.458	0.065
residuals	137	7.018	0.051		

Table 2: ANOVA table on cooperation (individual averages).

but was also economically efficient. In this respect, the second best treatment was *No-Neg* (67%), followed by *Ex-Null* (60%), *No-Pos* (59%), *No-Null* and *Ex-Neg* (both 58%).

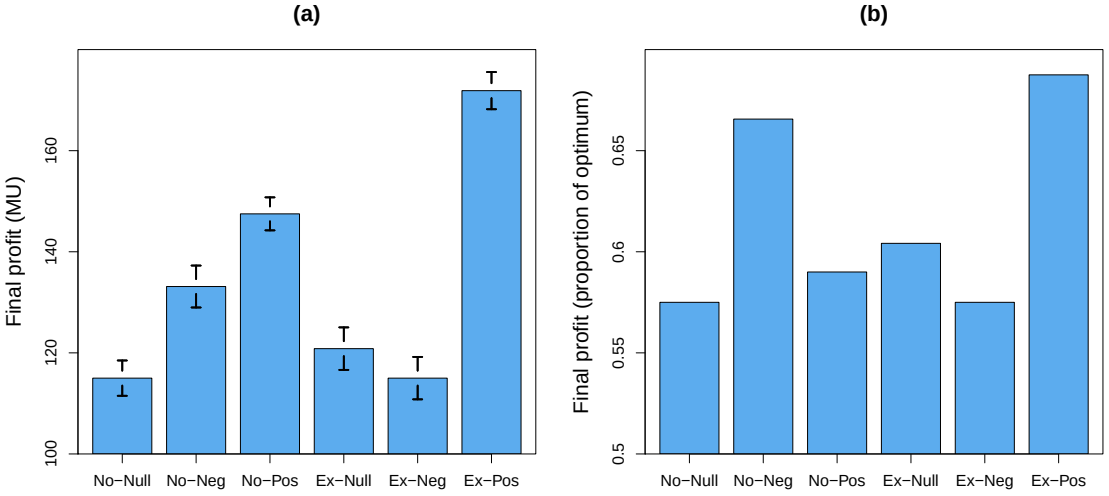

Figure 3: (a) Average final profit per treatment with standard error bars. (b) Total profit per treatment as proportion of the optimum.

A further confirmation of the positive interaction effect between exit and incentives on earnings is in the analysis of variance presented in Table 3. Besides the expected significant effect of rewards, it is worth noting that there were also significant interaction effects between exit and rewards and between exit and punishment, the latter leading to lower average profits.

6 Discussion

The fact that individuals are sensitive to the magnitude of rewards and punishments and, when rewards and punishment are considerable, cooperation tends to proliferate is generally ackowledged. However, institutions do not always succeed in providing sufficient incentives to avoid free-riding temptations. This motivated us to examine a situation where a sanctioning system existed but not sufficiently to change the dominant strategy of the players. Following recent theoretical investigation, we added voluntary partici-

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
exit	1	584.03	584.03	1.64	0.202
rewarding	1	47920.92	47920.92	134.75	0.000
punishment	1	906.51	906.51	2.55	0.113
exit × rewarding	1	7452.17	7452.17	20.96	0.000
exit × punishment	1	3444.01	3444.01	9.68	0.002
residuals	138	49075.03	355.62		

Table 3: ANOVA table on participants' final profits.

pation in the the game as a second factor potentially able to increase cooperation levels (Batali and Kitcher 1995; Hauert et al. 2002; Sasaki et al. 2007), notably in interaction with institutional actions (Sasaki et al. 2012).

Our experiment confirmed the strength of punishment in motivating cooperation, while rewards led only to small improvements if individually considered. We found that, although sanctions were theoretically insufficient to alter subjects' rational preferences, *No-Neg* determined a prevalence of cooperation. This is in contrast with the idea of a detrimental effect of sanctions on human altruism (Fehr and Rockenbach 2003) and, more generally, with the idea that monetary incentives can crowd-out intrinsic motivations (Bowles 2008; Frey and Jegen 2001).

Indeed, we found that even the imposition of small fines led to a significant increase in cooperation. A possible explanation is that fines might have a signalling function in highlighting misbehaviour. Although our experimental instruction were abstract and simplified, by using "incentives" and "disincentives" instead of "rewards" and "fines", it is possible that by penalizing noncooperative action subjects framed the game as a moral decision and were induced to cooperation more than what rationally expected, even if this led to lower earnings.

In contrast with our third hypothesis and with certain previous studies (Orbell and Dawes 1993; Semmann et al. 2003), voluntary participation did not increase cooperation if individually considered. This was due to the fact that, being participation a volutary decision, intending defectors were not motivated to opt out and, therefore, this did not provide room for cooperative equilibrium. Even if this happens in many real-world situations, an interesting extension of our study could be to introduce a participation cost or, conversely, a fixed reward for non-participation.

While voluntary participation did not improve the situation by itself, it determined a significant increase of cooperation when coupled with rewards. This finding is consistent with Sasaki et al. (2012), who argued that a positive interplay between institutional incentives and voluntary participation could exist, although we could not support their hypothesis on the superiority of punishment over rewards. Note that this difference may be due to the fact that, in order to simplify the game structure in a set of understandable instructions, we introduced fixed incentives and assumed that rewards and penalties did not depend on the number of cooperators and defectors in the population.

It is worth noting that the considerable cooperation level in *Ex-Pos* was not due to intending defectors choosing not to participate in the game. Indeed, these players had no rational incentive to abstain from playing and actually chose to exit only in a few cases. This could be explained in terms of a "frame effect" (Tenbrusel and Messick 1999): com-

bined with exit, not only did the presence of rewards induce subjects to expect that only well-intentioned subjects would have participated, but, more importantly, this induced subjects to attach to cooperative strategies higher expected payoffs and predict more cooperation from other subjects.

When considering the aggregate benefit of players, *Ex-Pos* was the treatment with the highest earnings, both in absolute terms and considering the extra money provided by the institution itself. This result suggests that institutions and organizations could improve their performance by setting up rewards while giving individuals the chance of voluntarily choosing whether to participate. Indeed, this could have a frame effect coherent with the positive nature of the incentive and induce individuals to expect more good behaviour from others.

To sum up, although weakly significant in itself, voluntary participation led to an increase of cooperation in commons dilemmas when combined with institutional enforcement. Obviously, in the case of real organizations and institutions, there is no perfect monitoring and some free-riding behaviour may remain unpunished. In this respect, an interesting extension of our work would be to consider monitoring costs and/or asymmetry of information such that subjects could expect not to be caught out with a given probability. This could lower the relative good performance of punishment, whereas the negative effect could less considerable for rewards. However, our results showed that in situations where there is little room for good behaviour, even weakly institutionally built-in positive signals for social interaction (i.e., little rewards and voluntariness) can modify the fate of the tragedy of the commons.

Acknowledgements

We gratefully acknowledge the help provided by Karl Sigmund in the design of the experiment and by Niccolò Casnici in its implementation. We would also like to thank three anonymous journal referees for important comments and suggestions.

References

Agrawal, A. (2002). Common resources and institutional sustainability. In E. Ostrom, T. Dietz, N. Dolšak, P. C. Stern, S. Stonich, and E. U. Weber (Eds.), *The Drama of the Commons*, pp. 41–85. Washington DC: National Academy Press.

Andreoni, J. and R. Croson (2008). Partners versus strangers: Random rematching in public goods experiments. In C. R. Plott and V. L. Smith (Eds.), *The Handbook of Experimental Economics Results*, Volume 1, pp. 776–783. North-Holland: Elsevier.

Andreoni, J., W. Harbaugh, and L. Vesterlund (2003). The carrot and the stick: Rewards, punishment and cooperation. *American Economic Review 93*(3), 893–902.

Barrera, D. and V. Buskens (2009). Third-party effects on trust in an embedded investment game. In K. Cook, C. Snijders, V. Buskens, and C. Cheshire (Eds.), *eTrust: Forming Relationships in the Online World*, pp. 37–72. New York: Russell Sage.

Batali, J. and P. Kitcher (1995). Evolution of altriusm in optional and compulsory games. *Journal of Theoretical Biology 175*(2), 161–171.

Boero, R., G. Bravo, M. Castellani, and F. Squazzoni (2009). Reputational cues in repeated trust games. *Journal of Socio-Economics 38*(6), 871–877.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. *Science 320*, 1605–1609.

Burnham, T. C. and B. Hare (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature 18*, 88–108.

Buskens, V., W. Raub, and J. van der Veer (2010). Trust in triads: An experimental study. *Social Networks 32*, 301–312.

Corten, R. and V. Buskens (2010). Co-evolution of conventions and networks: An experimental study. *Social Networks 32*, 4–15.

Dari-Mattiacci, G. and G. D. Geest (2010). Carrots, sticks, and the multiplication effect. *Journal of Law, Economics, and Organization 26*(2), 365–384.

de Quervain, D. J.-F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr (2004). The neural basis of altruistic punishment. *Science 305*, 1254–1258.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*, 980–994.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature 415*, 137–140.

Fehr, E. and B. Rockenbach (2003). Detrimental effects of sanctions on human altruism. *Nature 422*, 137–140.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 171–178.

Frey, B. S. and R. Jegen (2001). Motivation crowding theory. *Journal of Economic Surveys 15*, 589–611.

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., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science 322*, 1510.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior 24*, 153–172.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und Wissenschaftliches Rechnen 2003*, pp. 79–93. Göttingen: Ges. für Wiss. Datenverarbeitung.

Hardin, G. (1968). The tragedy of the commons. *Science 162*(3859), 1243–1248.

Hauert, C., S. De Monte, J. Hofbauer, and K. Sigmund (2002). Volunteering as red queen mechanism for cooperation in public goods games. *Science 296*, 1129–1132.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*, 1362–1367.

Isaac, M. R., K. McCue, and C. Plott (1985). Public good provision in an experimental environment. *Journal of Public Economics 26*, 51–74.

Janssen, M. A., R. Holahan, A. Lee, and E. Ostrom (2010). Lab experiments for the study of social-ecological systems. *Science 328*, 613–617.

Kahneman, D. and A. Tversky (Eds.) (2000). *Choices, Values and Frames*. Cambridge: Cambridge University Press / Russell Sage Foundation.

Kosfeld, M., A. Okada, and A. Riedl (2009). Institution formation in public goods games. *The American Economic Review 99*(4), 1335–1355.

Ledyard, J. (1995). Public goods experiments. In J. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*, pp. 111–194. Princeton: Princeton University Press.

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.

North, D. C. (2005). *Understanding the Process of Economic Change*. Princeton, NJ: Princeton

University Press.

Okhuysen, G. A. and K. M. Eisenhardt (2002). Integrating knowledge in groups: How formal interventions enable flexibility. *Organization Science 13*(4), 370–386.

Orbell, J. M. and R. M. Dawes (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American Sociological Review 58*(6), 787–800.

Ostrom, E. (1990). *Governing the Commons. The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. *American Political Science Review 92*, 1–22.

Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton: Princeton University Press.

Ostrom, E. (2006). The value-added of laboratory experiments for the study of institutions and common-pool resources. *Journal of Economic Behavior & Organization 61*, 149–163.

Ostrom, E., R. Gardner, and J. Walker (1994). *Rules, Games, and Common-Pool Resources*. Ann Arbor: The University of Michigan Press.

Ostrom, E., J. Walker, and R. Gardner (1992). Convenants with and without a sword: Self-governance is possible. *American Political Science Review 86*, 404–417.

Rand, D. G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak (2009). Positive interactions promote public cooperation. *Science 325*, 1272–1275.

Richerson, P. J. and R. Boyd (2001). The biology of commitment to groups: A tribal instincts hypothesis. In R. Nesse (Ed.), *Evolution and the Capacity for Commitment*, pp. 186–220. New York: Russell Sage Foundation.

Sasaki, T., A. Brännström, U. Dieckmann, and K. Sigmund (2012). The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proceedings of the National Academy of Sciences 109*(4), 1165–1169.

Sasaki, T., I. Okada, and T. Unemi (2007). Probabilistic participation in public goods games. *Proceedings of the Royal Society B: Biological Sciences 274*(1625), 2639–2642.

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.

Semmann, D., H.-J. Krambeck, and M. Milinski (2003). Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature 425*, 390–393.

Spicer, M. W. (1985). A public choice approach to motivating people in bureaucratic organizations. *Academy of Management Journal 10*(3), 518–526.

Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *The Review of Economic Studies 77*(4), 1540–1566.

Tenbrusel, A. E. and D. M. Messick (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly 44*, 684–707.

Walker, J., R. Gardner, and E. Ostrom (1990). Rent dissipation in a limited access common-pool resource: Experimental evidence. *Journal of Environmental Economics and Management 19*, 203–211.

Walker, J. M., R. Gardner, A. Herr, and E. Ostrom (2000). Collective choice in the commons: Experimental results on proposed allocation rules and votes. *The Economic Journal 110*, 212–234.

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.

Whitmeyer, J. M. (2002). The compliance you need for a cost you can afford: How to use individual and collective sanctions? *Social Science Research 31*, 630–652.