Theory and Decision
https://doi.org/10.1007/s11238-023-09929-9

Voluntary play increases cooperation in the presence of punishment: a lab in the field experiment

Francesca Pancotto[1] · Simone Righi[2] · Károly Takács[3,4]

Accepted: 2 April 2023

Abstract
Problems of cooperation have often been simplified as the choice between defection and cooperation, although in many empirical situations it is also possible to walk away from the interaction. We present the results of two lab-in-the-field experiments with a diverse pool of subjects who play optional and compulsory public goods games both with and without punishment. We find that the most important institution to foster cooperation is punishment, which is more effective in a compulsory game. In contrast to Rand and Nowak (Nat Commun 2(1):1–7, 2011), we find that loners are not responsible for anti-social punishment, which is mostly imputable to low-contributors (free-riders). Loners neither totally free-ride nor they significantly punish cooperators (or other types of players): they simply avoid all forms of participation whenever possible.

Keywords Cooperation · Optional public goods game · Exit · Punishment · Lab in the field experiment

1 Introduction

The studies of mechanisms and conditions that support cooperation among unrelated individuals are foundational for both social scientists and researchers of animal behaviour (Bicchieri, 1997; Riolo et al., 2001). Social dilemma games, in which individuals face a binary choice between a cooperative and a non-cooperative action, are among the most important theoretical and experimental tools through which cooperation is studied (Trivers, 1971). Many real-life situations, however, cannot be simplified to a mere choice between cooperating and not cooperating. Individuals often can also opt-out from interactions, giving up both potential benefits of cooperation and potential gains from free-riding. For example, a person can decide to

Extended author information available on the last page of the article

Published online: 13 May 2023

not to join a business alliance (Seale et al., 2006), to quit a job if she does not like the work environment, to abstain from attending a game of a sports team or the premiere performance of her orchestra. In all these cases, the person instead of opting-out could have joined and withhold the effort associated with the creation of the public good of the group (e.g., the objective of the work group, the performance of the sport team or orchestra), thus acting as a free-rider. Similarly, abstaining from a work in a company or from a research group in academia, are examples in which an employee does not cooperate anymore with others in his group, while also not free-riding. Participation in an international political or military alliance (such as the European Union or NATO) is also a possible example of cooperation with the possibility of opt-out. Being a participating member who contributes to the joint budget corresponds to the role of the cooperator, while a free-rider country is one who reaps the benefits of the alliance, without contributing to its balance (e.g., failing to invest sufficient resources in the defense sector for a NATO member). Finally, a country can decide to stay out of the alliance altogether, or to exit it, giving up both the benefits from the alliance (e.g., protection, free trade) and its participation costs (e.g., defense investments, contributions to the common budget). We will label the latter kind of actors as loners.

The scientific literature has conceptualized not taking part in a social dilemma interaction in multiple ways: as exit after play (Schuessler, 1989; Phelan et al., 2005; Vanberg & Congleton, 1992), as walking away (Aktipis, 2004), as opting-out (Orbell & Dawes, 1993; Batali and Kitcher, 1995; Hauert et al., 2002a, b), and—when opting out is not voluntary—as exclusion (Liu et al., 2018, 2019). In all these cases, cooperation is enhanced by the possibility of non-participating through various mechanisms such as: the possibility to exit the game after defection of the interaction partner and the threat it imposes on play (Schuessler 1989), the round-robin elimination of opportunistic counterparts due to their lower overall performance in repeated games (Phelan et al., 2005), the positive assortment of cooperation strategies favoured by the exit and exclusion mechanisms (Aktipis, 2004; Liu et al., 2018, 2019), and its interaction with reputational mechanisms (Podder et al., 2021). Relatedly, when the population of loners is large enough, cooperating survives evolutionary pressure (Batali and Kitcher, 1995) leading some authors to suggest that a plausible scenario to start cooperation is a world dominated by loners, not by defectors (Castro & Toro, 2008). Further, opting-out in a public goods game solves the problem of cooperation by undermining the free-riding strategy of defectors that cannot exploit loners (Hauert et al., 2002a, 2002b). In any case, the relevant parameter for the outcome is the outside option associated with refusing to play (Macy & Skvoretz, 1998).

Despite the many theoretical arguments proposing that voluntary participation may stimulate cooperation, only a few empirical studies have examined the issue (Orbell & Dawes, 1993; Nosenzo and Tufano, 2017; Ehrhart et al., 1999; Hauk 2003; My & Chalvignac, 2010). These studies compare the effects of entry and exit options on cooperation, using a public goods game experiment. They show that the opt-out or exit option is capable of sustaining cooperation through the value of the threat, while the entry strategy is not as effective (Nosenzo & Tufano, 2017). Thus, these results highlight that exit and opting-out are efficient institutions for enhancing cooperation when they are credible threats, i.e. when there are other interactions or benefits outside the current interactions. In absence of an outside option, there is no evidence that opting out can enhance cooperation.

In this paper, we highlight that the availability of another institution, punishment, sheds light on the motivations behind the choice to opt-out. The focus on the combination of opt-out and punishment differentiates our contribution from experimental exercises in which there is a voluntary choice in the selection of a contribution out of different options (Chakravarty & Fonseca, 2017) and from those on endogenous group formation (Brekke et al., 2011; Page et al., 2005). In fact, none of the previous contributions explored the interaction between opting-out and punishment behaviour. Several real-life situations that can be conceptualized as optional public good games, involve also the possibility of punishment. Returning to the example of economic and political alliances, the 'Brexit' example (with the UK leaving the European Union) makes it clear that opting out can entail the possibility of being punished for this action by participants, which can decide to impose sanctions and trade restriction to the party opting out. It is also noticeable how the incentives to punish vary significantly when we consider the different positions toward the UK of net-contributors to the Union and net-receivers of EU funds.

Considering the interaction between punishment and opt-out, it is not only important for finding out the true motivational background behind the latter, but allows also to gain a deeper understanding of the conditions under which such interaction builds efficient institutions for solving social dilemmas. While post-interaction punishment alone has been demonstrated to provide a powerful solution to social dilemmas, both by theoretical (Helbing et al., 2010; Chen et al., 2015; Fang et al., 2019) and experimental work (Fehr and Gächter, 2000; Boyd et al., 2003), it remains an open question how opting-out impacts the effectiveness of the punishment institution. When punishment is interacted with opting-out, the set of possible punishment strategies is significantly increased. Outcomes might differ depending on whether only certain punishment strategies are allowed—e.g., when players can punish defectors but not cooperators (Fowler, 2005) or whether all strategies can be punished (Brandt et al., 2006; Rand & Nowak, 2011). While the first case leads to high cooperation, the second generates cyclic behaviours—confirmed in experiments (Semmann et al., 2003)—that do not allow any prediction concerning the sustainability of cooperation. Similarly, the success of cooperators who also decide to punish defectors depends on the presence of an opt-out strategy which provides larger payoffs than the the payoff of defectors who are punished (Hauert et al., 2007).

In the present paper, we study how the introduction of punishment impacts the ability of opting-out to induce cooperation. We run two experiments, where we study the introduction of punishment in a public goods game with and without the possibility of opting out from participation. We select a sample from the general local population of two Italian regions aimed at maximizing diversity (detailed discussion of the sample is reported in the Sect. 2.2). A similar design to ours is the one devised by Rand & Nowak (2011), which tests punishment behaviour in an optional public goods game in a between-subject design, in order to study whether loners engage in anti-social punishment when forced to participate (Herrmann et al., 2008). Similarly to us, Nosenzo and Tufano (2017) test the implications of the introduction of institutional variations of participation on cooperation. In this regard, our contribution is to explore the role of punishment on the ability of the opting-out mechanism to generate cooperation.

Given the presence of voluntary punishment opportunities in empirical social dilemma situations characterized by the opt-out option, the interaction between these two elements is an important path to follow. It is important to study the two institutions together, in order to assess whether the effectiveness of punishment in sustaining cooperation, depends on whether the social dilemma is characterized by optional or compulsory participation. With our design, we are also capable to pin down the reasons why the two institutions may be more effective together or separately. In fact, we can track down the behaviour of those who always participate in games and those who opt-out whenever possible (loners), both with and without the presence of punishment, and assess how their contributions differs for alternative institutional settings.

We run a within-subject design with the objective of comparing the behaviour of the same participant in different institutional conditions, i.e. with and without the option to participate, with and without punishment. This design, similar to Rand & Nowak (2011), allows to identify those who decide to not participate and to compare their behavior in the optional game to one in the mandatory game, both without and with punishment. As we propose different institutions to the same pool of subjects, clearly there is a between-game effect, because the same participants play all the games one after the other. This between-game learning, is functional to the objectives of the research: once participants have expressed their preference about participation in the optional game and then are forced to participate in the compulsory game, we can study their behavior in the latter, knowing that they would rather be out of the game if given the opportunity. Similarly, when punishment is introduced, we can track down the optional vs compulsory game effect interacted with the presence of the punishment institution, knowing that participants have experienced the same choice situation without punishment, and verifying how they adopt or reject punishment as a tool to induce cooperation. The final contribution of our work is the ability to explore also punishment behaviour of participants and loners, and provides a test of Rand & Nowak (2011) hypotheses on the behaviour of loners as punishers of cooperators, with a diverse sample and a lab-in-the-field in-presence experimental setting.

1.1 Hypotheses

Based on the literature discussed so far we are able to formulate five hypotheses concerning the relationships between opt-out and punishment. The literature on the effect of the option to participate on cooperation (Orbell & Dawes, 1993; Nosenzo and Tufano, 2017; Hauk, 2003; My & Chalvignac, 2010), suggests that in a public goods game, the option to exit the interaction is an effective mechanism to sustain cooperation, when the outside option is a credible threat. This led us to hypotesize that,

Hypotheses 1 Optional play induces higher cooperation than compulsory play in a public goods game where punishment is not allowed.

While there is no literature using punishment in an experiment with the optional public goods game, building on the corresponding literature on the public goods game with compulsory participation which suggests that cooperators are often able to sustain high contributions to the public good through costly monetary punishment of free-riders (Fehr and Gächter, 2000; Chaudhuri, 2011), we expect that punishment would exert a significant positive effect on cooperation. Thus, we hypothesize that:

Hypotheses 2 Introducing the punishment institution increases cooperation in public good games, both with compulsory and with optional participation.

The literature does not provide any reference to predict whether compulsory or optional play would generate higher or lower contributions, when punishment is present. However, our intuition is that punishment is more effective when compulsory play is present. Indeed, in a repeated interaction, free riders can use the option to exit to avoid being punished as well as loners opt out of participation if they are not willing to cooperate.

Hypotheses 3 In the presence of the punishment institution, compulsory play results in higher cooperation than optional play.

Finally, building on Rand & Nowak (2011), which one of our experiments largely replicates in a more diverse (see the next section for details) sample and with a lab-in-the-field setup, we will test two further hypotheses [corresponding respectively to HP1 and HP2 in Rand & Nowak (2011)]. First, we hypothesize that:

Hypotheses 4 Loners are low contributors when forced to participate in a public goods game.

Previous literature has clarified the role of high and low contributors for social and anti-social punishment (Herrmann et al., 2008). While high contributors likely punish low contributors (social punishment), low contributors can be perverse punishers, punishing high contributors (with a degree of heterogeneity, see Herrmann et al. (2008)).

According to Rand & Nowak (2011), loners are responsible for the punishment of high contributors, as a form of defense against their imposition of unwanted participation to the game. As a consequence, we expect two behaviors: first, that low contributors in compulsory games will opt for the loner's payoff if given the chance, and this is the reason in support of our HP4. Second, that loners can become anti-social punishers when forced to participate in a public goods game, which is our HP5:

Hypotheses 5 Loners are anti-social punishers when forced to participate in a public goods game.

2 Materials and methods

In order to test our hypotheses, we run two experiments on a total of $N = 236$ different subjects, for a total of 16 sessions, with every individual allowed to participate in only one session. We will call the first experiment E1 (comprising 176 subjects in 11 sessions) and the second E2 (60 subjects in 5 sessions).

The number of participants was 16 for each experimental session. In some cases, due to last minute drop-out, we run sessions with 12 subjects. Average session time was 1 h, and payoffs were expressed in experimental points (tokens), with each token corresponding to 0.04 €. The average payments per person was around 15 Euros, plus a show up fee of 5 euros.

2.1 Recruitment and sample selection

The problem of cooperation is intrinsically linked with the problem of social norms (Bicchieri, 2005; Bigoni et al., 2016) and involves society as a whole. For this reason, we selected a sample from the general local population at the time of the experiment. However, due to limitations in the pool of available potential participants, the actual sample over-represents females, students and younger individuals and under-represents males, retired, inactive, and older people, with respect to the Italian population as photographed by the 2011 census. Table 1 reports the sample composition and the discrepancy with respect to the underlying population. Despite these limitations, a broad spectrum of the population is covered by the sample, unlike in prevalent lab experiments with university students or online experiments where participants are more difficult to monitor.

The experiment was conducted with the use of a mobile laboratory for experimental economics. The experimental software was developed in Python using the o-Tree platform (Chen et al., 2016). This software platform was chosen as it allowed running experiments on devices with touchscreens (such as tablets) and with a web-based graphical user interface. These are more likely to be familiar to the wider population, which does not necessarily possesses computer proficiency. Accordingly, participants to all sessions assessed positively the easiness of use of both o-Tree and the tablets (see Table S18 in Supplementary Material).

Experimental sessions were run in the Italian provinces of Modena (in the municipalities of Vignola and Mirandola), Reggio Emilia (in the main town of the

Table 1 Sample composition and discrepancy with respect to national population

	Italy	Sample		Italy	Sample
Age class			Gender		
18–25	10%	37%	Male	49%	38%
26–35	15%	21%	Female	51%	63%
36–45	19%	12%	Occupational status		
46–55	18%	14%	Employed	57%	49%
56–65	15%	10%	Retired/inactive	31%	17%
65 +	25%	6%	Students	12%	34%

province), Macerata (in the main town of the province), Fiastra (Province of Macerata), Petritoli (Province of Fermo) between November 2015 and May 2018. In Vignola, Petritoli and Mirandola the experimental sessions were organized in the municipalities' council chambers while in Reggio Emilia, the experimental sessions were organized at the university. Sessions in Macerata were run at the Department of Economics of the University; sessions in Fiastra were run on a container used by citizens as recovery at the time of the emergency, made available by a local association. All venues used for the experiment were accessible by car and public transport. The general characteristics of the towns selected are described in Table S19 of Supplementary Material.

The sample was recruited (for each location) from the general population, with the aim of maximizing the diversity of the sample. In order to reach potential volunteers, 1000 letters were sent to random families selected from the lists of residents in the municipalities involved. Moreover, written advertisements (flyers and posters) were posted in a large number of restaurants, bars, shops. Further general diffusion was obtained with advertisements through the municipalities newsletter and Facebook pages. Finally, about 4000 ex-students from all faculties of the University of Modena and Reggio Emilia (graduated from 2009 to 2015) were contacted through email, inviting them to spread the information about the experiment. Potential volunteers were invited to register for the experiment either through a web–form, by email or by phone. Registered individuals were then randomly assigned to one of the sessions of the experiments realized in their municipality. All sessions were run on Saturdays in order to promote a wider and more diverse participation.

The pool of candidates registered for the experiments were selected for sessions imposing two restrictions. First, the candidate had to be 18 years or older at the time of the experiment (this restriction was required for legal reasons). Second, only residents of either the administrative municipalities in which the session was run, or of the neighboring ones, could participate in an experiment run there. Locations of the sessions, dates and times are reported in Table S18.

The experiments were conducted in accordance with regulations and relevant guidelines for experiments with human subjects of the REBEL (Reggio Emilia Behavioural Economics Laboratory) at the University of Modena and Reggio Emilia and approved by the REBEL ethics committee.

2.1.1 Experimental procedures

Upon arrival at the experimental session, subjects were registered and assigned a seat where they were given an informed consent, a privacy consent, and data release form to read and sign. Participants were informed that they were allowed to leave at any moment and obtain the show up fee, but nobody left the sessions. Participants were made aware of the fact that oral communication was forbidden during the experimental sessions. Moreover, mobile cubicles were used to make visual contact among participants impossible.

All instructions were read aloud by one (the same for all sessions) experienced experimenter. Moreover, the instructions for the current game were available at the bottom of the screen at any time during the associated task.

For each part of the experiment, instructions for the task were read aloud by the experimenter. Before moving on to the payoff relevant decisions, subjects were then asked to respond to 3–4 control questions, which would not affect the game's earnings. Given the complex nature of the recruiting procedure, we did not exclude subjects that did not respond correctly to the answers, instead we gave clear feedback on the control questions to the participants. After answering to the control questions, in the next page of the software, participants were provided with a calculated explanation in order to help them to fully understand the game structure. Moreover, we collected data related to the performance of the participants in the game and used the variable as a control in the analysis.

At the end of the sessions of both experiments, we run a guessing game and a dictator game (four players) as controls for strategic reasoning (cognitive ability) and individual pro-sociality (unrelated to community norms of cooperation). Results of these games are not analysed in this paper and are reported in Pancotto and Righi (2021). Finally, participants were asked to fill up a survey with demographics data and provided feedback about all the games played, together with the conversion of experimental tokens in Euros. Payments were made privately in cash at the end of the session.

2.2 Experimental design

2.2.1 The public goods game

All tasks discussed in this paper were variations of the same basic discrete version of the public good game, each featuring different institutional variations. In each round of the game, session participants were divided in groups of four. Each group member received an endowment of $e = 40$ experimental tokens and had to decide (simultaneously with other group members) how much of its endowment to invest in a common project, choosing a discrete contribution level ($c_i \in 0, 10, 20, 30, 40$) and knowing that the residual ($e - c_i$) would remain in their private account. The tokens shared by the individual members were summed up, doubled and shared equally among group members. Hence, individual earnings were determined as follows:

$$\pi_i = e - c_i + \alpha \sum_{j=0}^{N} c_j,$$

where α is the marginal per-capita return (MPCR) of the public good. Free-riding is the dominant strategy, for rational self-interested individuals, when MPCR is above $1/N$ and below 1. Social welfare is instead maximised when everyone contributes the whole endowment. The discrete contributions framework was chosen with the aim to facilitate calculations and understanding of the game for the participants.

2.2.2 Optional participation

In rounds of the game where participation is optional, by opting out, the individual reward (outside option) is the initial endowment ($c_i = 40$), while earnings from the public good are divided, in equal parts, only among participants.

As a consequence of this choice, the MPCR decreases as a function of the number of participants, given that the multiplication factor remains fixed at two, regardless of the number of participants. For $N = 0$, nobody participates and the PGG does not take place. Notably, the case of nobody participating to the PGG and the case in which all participants contribute $c_i = 0$ are ex-post payoff equivalent for a group member. Indeed, in both cases loners and free-riders (zero contributors) obtain the same payoff, $c_i = 40$. However, before knowing how many will contribute to the PGG, free-riders may expect a positive return from the PGG, while the loners' payoff is fixed and independent of others' decisions. This consideration highlights how loners might lack trust towards other group members. Indeed, from an individual payoff point of view, the Nash equilibrium implies that it would always be more convenient to participate to the public goods game and free ride, rather than to stay out of it entirely.

2.2.3 Punishment

In rounds of the game in which the punishment is available, subjects can choose among three possible actions, A, B or C: choosing option A has no effect on either player; choosing option B causes a player to lose 4 points while the target player loses 12 points; and choosing option C causes a player to lose 8 points while the target player loses 24 points. This punishment structure correspond to a 1:3 punishment technology.

Punishment is carried out with the strategy method, originally proposed by Selten et al. (2003). Indeed, the participant must indicate which action (A, B, C) they would take towards group members choosing each possible contribution level (0, 10, 20, 30, 40) and toward players opting-out of the game (in rounds in which the option is available to participants). The computer would then assign, randomly, pairs of participants in each group, and calculate the relative punishment levels, depending on the choices actually implemented by the pair. For example, agent A can be randomly assigned to punish agent B from his group and being punished according to the decisions of agents C (with agent B and C that can be both equal or different) from his group of four. The punishment of agent A toward agent B will be determined by the strategy method's choice of agent A for the level of contribution chosen by agent B. The punishment of A for his contribution level will be the one chosen by C for that contribution level. Subjects are allowed to condition their punishment choice on the other player's contribution/participation during that round as they can observe the amount contributed by the other players in the group. However, - as noted - they cannot directly target their punishment at a specific subject, but only at a level of contribution. It is worth noticing that, thanks to this punishment structure, subjects are always allowed to decide not to punish, which means that the punishment decision is always optional. Subjects were informed of this procedure and payment structure.

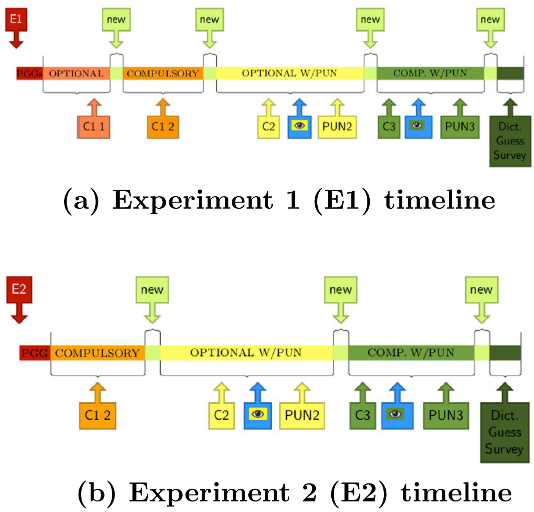
The two experiments that we run follow the timelines reported respectively in Fig. 1a for E1 and Fig. 1b for E2. In all cases, the public goods games were run with the parameters reported above. The translation of the experimental instructions is provided in Section S4 of Supplementary Material.

2.2.4 Experiment 1

The first experiment (E1) involved 176 participants and was composed of three tasks. In the first task, an Optional Public Goods Game (O-PGG) was followed by a Compulsory Public Goods Game (C-PGG). Between each round of the game, groups were reshuffled and participants were informed that they would play with new—different—group members. Indeed, in order to obtain a stranger matching protocol, every time the participants played a different PGG, groups were reshuffled (in Fig. 1a, b, the 'new' light-green sign indicates that there is a group reshuffling).

In the Optional game, after explaining the rules of the PGG, subjects could decide to exit the game, in exchange for a fixed amount of points ($c_i = 40$). This payoff is the loners' payoff in Rand & Nowak (2011) model, where a loner gets a payoff which is lower than the payoff of a cooperator in a group of cooperators. We use the choice to opt-out in this stage to identify loners. Then, subjects who chose to participate, decided their contribution level (labelled C11). After reshuffling, subjects were asked to make a compulsory contribution decisions (COMPULSORY in the Figure), with the same structure as in the preceding O-PGG, but without the possibility to opt-out. This contribution decision in the compulsory game is labelled C12.

After further re-matching, we run the second part of the experiment, consisting of the Optional PGG with Punishment (O-PGG-P). The O-PGG-P, was composed of three decisions: first subjects had to decide whether to participate in the game. Only



(a) Experiment 1 (E1) timeline

(b) Experiment 2 (E2) timeline

Fig. 1 Time line of experiments E1 and E2. C11: contribution to the optional PGG without punishment; C12: Contribution to the compulsory PGG without punishment; C2: contribution to the optional PGG with punishment; C3: contribution to the compulsory PGG with punishment; new: groups are reshuffled; eye symbol indicates that participants observe contributions in that particular stage those who agreed to participate made a contribution decision in the PGG (C2). Second, all subjects in the group observed the decisions of other group members, i.e. how much each group member contributed, but without precise identification of the single participant that contributed each amount (this is indicated with the eye symbol in the timeline). This is the only moment where subjects received a feedback in one round of the experiment. Subjects were informed of the possibility of punishing and being punished and of the process through which it was going to happen before they started taking decisions. In the third decision of the second part, all subjects went through a punishment stage, in which everyone had to express whether and how much they wanted to punish each possible contribution level, as well as the exit option (for a total of 6 different choices), at a cost for themselves. This decision is indicated as PUN2 in Fig. 1a. The punishment decisions that became payoff-relevant was selected randomly by the computer, as previously described. This punishment method allows to obtain information on behavioral response to all possible decisions of others and helps to identify the strategies participants follow. With our design, we will also be able to assess the responses of loners, who are similar to third party punishers (Fehr and Fischbacher, 2004; Bravo et al., 2015) that observe the society and judge behaviours without being directly involved in it, with the notable difference that the loners can themselves be punished.

Finally, in the third part of E1, we run a compulsory PGG with punishment (C-PGG-W/P) identical to the game of the second part, but without the possibility to stay out of the game (C3). Participants observed the contribution levels of group members and then decided about the punishment level (PUN3). This compulsory version of the game aims at studying the behaviour of loners when forced to participate in a game with punishment.

This experimental design reproduces closely the setup of Rand & Nowak (2011), with three notable differences. The first is that we run the experiment in a lab-in-the-field setting and not online. The second was that loners were allowed to punish others in the O-PGG-P. The third was that the same participants were asked to play both the games without and with punishment, in this order. As our main objective is to assess the impact of introducing the punishment institution on subject behaviours this change of design is essential to our objective. Indeed, through this design we are able to assess the impact of having a mandatory participation, both without and with punishment.

2.2.5 Experiment 2

The second experiment (E2), run with 60 subjects different from the previous ones, is identical to E1 except for the fact that the initial O-PGG was removed and the participants moved directly to the C-PGG stage. The remaining parts of the design were identical to E1.

Just as in E1, subjects were not informed about the outcome of first part of the game before running the second and the third part. They only observed the contribution levels of other group members before the punishment stage (the eye symbol in the time line of Fig. 1a, b), even though they could not exactly identify the specific group member who contributed a specific amount.

Table 2 Comparison of contributions levels between E1 and E2 (Mann–Whitney test for independent samples)

Avg. contributions	E1 + E2 $N = 236$	E1 $N = 176$	MW E1 vs E2	E2 $N = 60$
Overall	26.35	26.45	$p = 0.18$	25.97
C11	25.91	25.91	–	–
C12	25.23	25.39	$p = 0.31$	25.16
C2	26.41	27.15	$p = 0.041$	24.44
C3	27.62	27.44	$p = 0.45$	28.16

We stress that both in E1 and E2, participants were allowed to punish also those who opted-out (loners). Furthermore, in contrast to Rand & Nowak (2011)'s design, loners are also asked to make a punishment decision.

3 Results

3.1 Contributions under different institutions

We first report overall contribution rates, together with a Mann–Whitney test (MW) to test whether the results from the two experiments are comparable.

Results of the two experiments are indeed comparable as the difference between overall average contributions is not significant (E1 = 26.45 vs E2 = 25.97, MW=59832, p = 0.3776, two-sided, Table 2, first line), and contributions in the different parts of the game between E1 and E2 are non significantly different, except for the Optional game with punishment (C2), which is significantly larger in E1 than in E2 at 5% (MW, p = 0.041). A possible explanation of this significant difference is the small difference in the design of the two experiments. In E1, subjects play the optional game with punishment after having played the same game without punishment at the beginning of the experiment. By contrast, there is no such game in the beginning of E2. Subjects may be prone to some kind of learning in E1, whereas this is not the case in E2.

Therefore, having established the comparability of the data between the two experiments, we analyze pooled data.[1]

In Fig. 2, we plot the mean contributions of each stage, indicating with shapes the Participation rule (triangle for optional, circle for compulsory participation) and with color the institution (green for a stage without punishment, blue for one with punishment). First, by comparing C11 and C12, we observe that the difference is quite small: a Wilcoxon matched pair test of the difference in mean contribution between stage C11 and C12 confirms that this difference is not statistically significant (p = 0.397). So, we state that:

Finding 1 *In the optional play without punishment and in the compulsory game contributions do not differ. We cannot confirm HP 1.*

[1] We however report results for E1 and E2 separately in supplementary material, with qualitatively equivalent results.

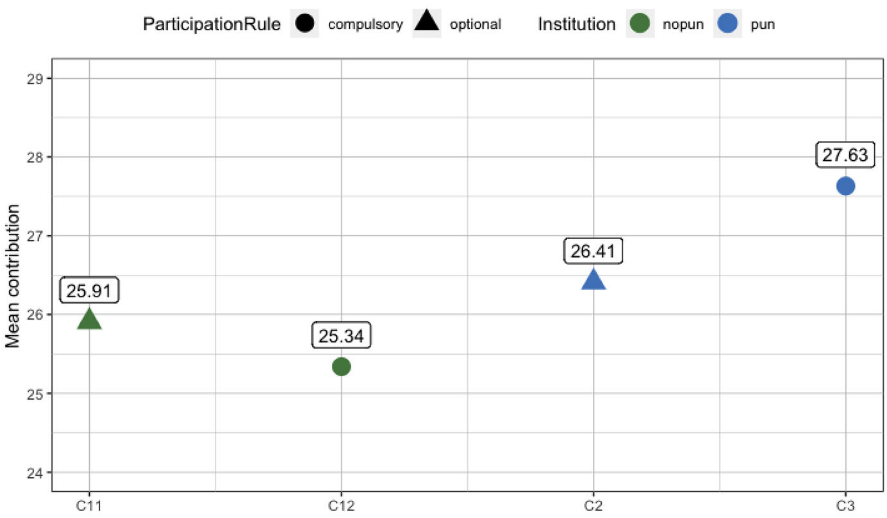


Fig. 2 Mean values of contributions for each institution and participation rule

Second, again in Fig. 2 we observe that contributions are higher in stages when the punishment institution is present (C2 and C3) with respect to stages without punishment (C11 and C12), independently from the participation rule. In fact, if we pool together contributions in optional and compulsory data without punishment, we obtain a value of 25.75 (mean contributions of C11 and C12), lower than the average contribution of stages with punishment (C2 and C3), which is 27.07. This difference is statistically significant using with a Wilcoxon matched pair test ($V = 5747$, and $p < 10^{-4}$).

We thus find support for the fact that:

Finding 2 *Contributions are significantly higher when the punishment institution is present. We confirm HP 2.*

Finally, we compare average contribution in the stage where both compulsory participation and punishment is present (C3) with all the other stages, with a Wilcoxon matched pair test [C11 (25.91) < C3 (27.62), $p = 0.006$, C12 (25.33) < C3 (27.62), $p = 0.002$, C2 (26.41) < C3 (27.62), $p = 0.011$] and find that:

Finding 3 *Contributions are the highest when punishment is possible and participation is compulsory. We confirm HP 3.*

In the next paragraph, we will explain these results by studying the decisions of participants and loners.

3.2 Individual cooperation

In this section, we explore further the results presented in Findings 2 and 3. To this end, we separate average contribution levels between participants and loners

Table 3 Average contributions of participants and loners

E1 + E2	All subjects	*N*. part	Participants (mean contr.)		Loners (mean contr.)	*N* loners
C11	25.91	164	25.91		–	12
			(↓ WX: $p = 0.39$)			
C12	25.33	–	25.73	(→ MW: $p = 0.090$)	20.83	
C2	26.41	198	26.41		(↓ WX: $p = 0.046$)	38
	–		(↓ WX: $p = 0.03$)			
C3	27.63	–	28.33	(→ MW: $p = 0.02$)	23.94	
E1	**All subjects**	***N* part.**	**Participants**		**Loners**	***N* loners**
C11	25.91	164	25.91	–	–	12
			(↓ WX: $p = 0.39$))			
C12	25.41		25.73	(→ MW: $p = 0.090$)	20.83	–
C2	27.15	144	27.15	–	(↓ WX: $p = 0.044$)	32
			(↓ WX: $p = 0.18$)			
C3	27.44		28.06	(→ MW: $p = 0.098$)	24.69	–
E2	**All subjects**	***N* part.**	**Participants**		**Loners**	***N* loners**
C1.2	25.16		25.74	(→ $p = 0.08$)	–	–
C2	24.44	54	24.44	–	–	6
			(↓ WX: 0.0028)	–	–	
C3	28.16		29.07	(→ MW: $p = 0.037$)	20.00	–

(Table 3), and look at how these two groups contribute under different institutional structures.

First, we compare loners and participants. As expected, loners contribute significantly less than participants in the compulsory game both without punishment (C1.2: participants = 25.73, loners = 20.83, MW, $p = 0.09$) and with punishment (C3: participants = 28.33, loners = 23.94, MW, $p = 0.02$). These results hold when looking at E1 and E2 separately (bottom panels of Table 3). This result supports our HP 3 and confirms the results of Rand & Nowak (2011):

Finding 4 *Loners contribute significantly less than participants in a compulsory game, showing a different willingness to cooperate. We confirm HP 4.*

Moving the attention to the behavior of participants, we first notice a statistically non-significant reduction in their contributions when comparing the optional (C11 = 25.91) and compulsory games (C12 = 25.33) without punishment. Such reduction is influenced not only by the expected lower contribution of loners but also by a small reduction in participants' contribution (25.91 versus 25.73).

When we consider instead the game with punishment, participants increase significantly their contributions when the compulsory play is introduced (C2 = 26.41 to C3 = 28.33, WX: $p = 0.03$). We speculate that the presence of punishment induces participants to increase their contribution as a response of higher trust in the expected cooperation of the group: if loners are expected to behave as free-riders when forced to participate, they could be still induced to cooperate at least partially by punishment, and this would suffice to support higher contributions from participants. This explanation is supported by survey response to a question on trust (formulated as in the World Value Survey), which becomes significant in explaining the contribution in compulsory play when punishment is present (Tables S1 and S2 in SI).

For what concerns loners, the data show that they are not free-riders and they contribute something even when they are not under the threat of punishment. Also in their case, punishment affects positively their willingness to cooperate as their average contribution increases from 20.83 to 23.94. Using a MW test of differences in distribution, we do not find a significant result (MW: 433.5, $p = 0.1409$): the test uses the mean value of participants who did not necessarily behave as loners in both stages. Using a Wilcoxon sign test with matched pairs instead, we find a significant difference (WX: $V = 54$, $p = 0.0467$). In this case the test matches data of people who can be identified as loners in both stages.

The WX test performs a matched pairs test which means that it considers only those participants that were loners in both cases: here we see a significant difference, which indicates that loners in the stage without punishment contribute less than in the stage with punishment.

Figure 3 reports the variation (delta) between the contribution levels in the two games as the difference between compulsory game and the corresponding optional game. It is important to stress that to calculate the increase of contributions in the

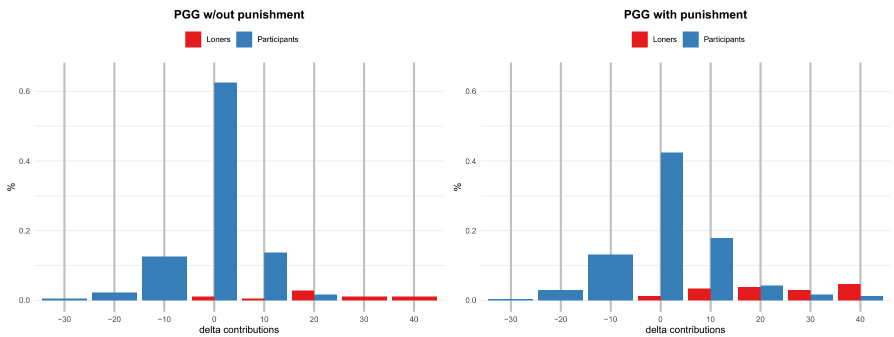


Fig. 3 A positive delta (+ 10 to + 40, right side of plots) indicates that an individual increased the contribution when moving from optional to the compulsory PGG. A negative delta (− 40 to − 10) indicates the opposite. Bars count how many participants and loners respectively change their mind as a total of all players. The bar on top of zero values, indicates no change. Left panel reports values for the delta in the no punishment case, right panel for the punishment

compulsory game for loners, we need to impose artificially the value of contributions in the optional game as zero. The left panel relates to the PGG without punishment, the right panel to the PGG with punishment. In the left panel, we see the highest bar corresponding to zero delta, which is also the highest bar in the game with punishment. This indicates that on average participants maintained more their zero contribution level in the game without punishment. The pattern of those who reduce their contributions are similar for participants in both figures, while we see higher increase in contributions for participants in the game with punishment, confirming our previous results: punishment makes participants and loners change their mind more often. Again, loners appear not to be high contributors, but the increase toward positive contribution levels is higher in the game with punishment.

3.3 Individual punishment behavior

We now report the punishment behavior in the optional and in the compulsory games in more detail. In order to test HP 5, we explore also the presence of anti-social punishment, defined by Herrmann et al. (2008) as the punishment in which the punisher contributes less than, or the same amount, as the target of his/her punishment. Other authors (Bochet et al., 2006), distinguish between 'normal' and 'perverse' punishment, where the latter is the one directed at the highest contributors (contribution levels equal to 30 and 40) or at someone who has contributed more than the average. In our setup, we use the first definition and call perverse the punishment toward contribution levels equal to 30 or 40. Normal is instead the punishment directed at contribution levels equal to 0 or 10. Finally, we study punishment of maximal contributors, i.e. those contributing exactly 40.

3.3.1 Optional game

In the optional game, the average punishment expenditure invested by players expressing each level of contribution and directed at all contribution levels and at loners is reported in Fig. 4 (numerical data in Tables S4 and S5 of Supplementary Material). The overall punishment expenditure is decreasing with the increase in the contribution level of the target. Moreover, high levels of punishment are directed at low contributors (left side of the figure) and are done by high contributors (green and light blue lines). Thus, we can conclude that—on average—punishment is of prevalently of the normal kind. However, we observe a positive level of punishment expenditure directed at people contributing 30 and 40, indicating the presence of some perverse punishment. Who is then responsible for perverse punishment? Looking at the punishment inflicted by low contributors (0 and 10 contributors), we notice that their punishment expenditure increases with the contribution of the target for levels above 20. Thus, punishment of maximal contributors is done mainly by low contributors (also called as free-riders). We conclude that:

Evidence 1 *We observe perverse (anti-social) punishment in the optional public goods game, where loners are not forced to participate. Low contributors engage in perverse punishment.*

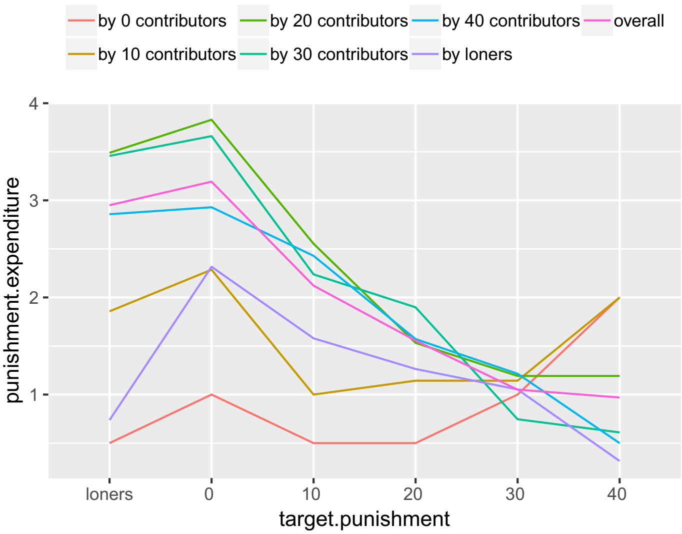


Fig. 4 Optional game. Average punishment expenditure by level of contribution directed at all targets, including loners

The punishment pattern is strikingly different for loners, who engage in very low levels of punishment toward maximal contributors (violet line in Fig. 4, precise values in Table S4 of Supplementary Material). This result is in sharp contrast with HP 5 and the findings of Rand & Nowak (2011). Moreover, the punishment of loners by all contribution levels (overall line in red), is affine to the one of low contributors. It is worth stressing that cooperators do not seem to differentiate between the behavior of loner and that of low contributors. The highest amount of punishment expenditure is directed towards loners by high contributors (see green lines in Fig. 4 on the left corresponding to loners). The punishment of loners by loners is also low (violet line is low at the left of the panel, at the x-tick mark corresponding to loners), helping to shed further light on their typical strategy. Loners support their decision by abstaining from punishing behaviours. Loners punish on average less than other participants (but more than 0-contributors), confirming that their main objective is to avoid interactions with the group. Summarizing findings concerning loners, we can state that:

Finding 5 *In the optional game, loners do not punish high contributors. We do not confirm HP 5.*

On the other hand,

Evidence 2 *Loners do not punish other loners. High contributors punish loners and low contributors similarly.*

Studying the results of multivariate logit models on the decision to punish maximal contributors and loners controlling for age, gender, and trust (Table 4), we

Table 4 Punishment of maximal contributors and loners: Optional game. (Logit Models)

	Dependent variable			
	40 (maximal)-contributors			Loners
	(1)	(2)	(3)	(4)
High contributors (C2)	− 1.276***			
	(0.452)			
Low contributors (C2)		1.333***		
		(0.457)		
Opt-out			− 1.402*	− 1.530***
			(0.773)	(0.483)
Age	0.020	0.017	0.015	0.005
	(0.015)	(0.015)	(0.014)	(0.010)
Gender	− 0.563	− 0.637	− 0.667*	− 0.257
	(0.426)	(0.426)	(0.398)	(0.296)
Trust	− 0.822**	− 0.781**	− 0.903**	0.194
	(0.391)	(0.397)	(0.369)	(0.265)
E1	0.499	0.121	0.208	− 0.841**
	(0.505)	(0.490)	(0.468)	(0.330)
Constant	− 0.199	− 0.704	− 0.182	0.493
	(0.991)	(1.015)	(0.931)	(0.723)
Observations	185	185	223	223
Akaike Inf. Crit	159.848	160.217	182.852	290.658

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

find that high contributors do not punish their own type, while low contributors are responsible for more punishment of high contributors. These results are robust when we adopt as dependent variable the punishment expenditure directed at each target contribution and at loners, and as regressors, contribution level of the punisher, both with and without controls (Tables S6–S11). Further, from Table 4, we confirm that the decision to opt-out is significantly and negatively related to the punishment of maximal contributors, i.e., that loners are not responsible for perverse punishment (Evidence 1). Finally, the opt-out decision is significantly and negatively related to the punishment of loners, which confirms the evidence that loners do not punish other loners (Evidence 2).

3.3.2 Compulsory game

In the compulsory game, punishment decreases by the contribution level of the target (Fig. 5). For high and medium contributors, the higher the contribution of the punisher, the lower the punishment of high contributors, i.e., punishment is prevalently of the normal type. However, the opposite happens for low contributors

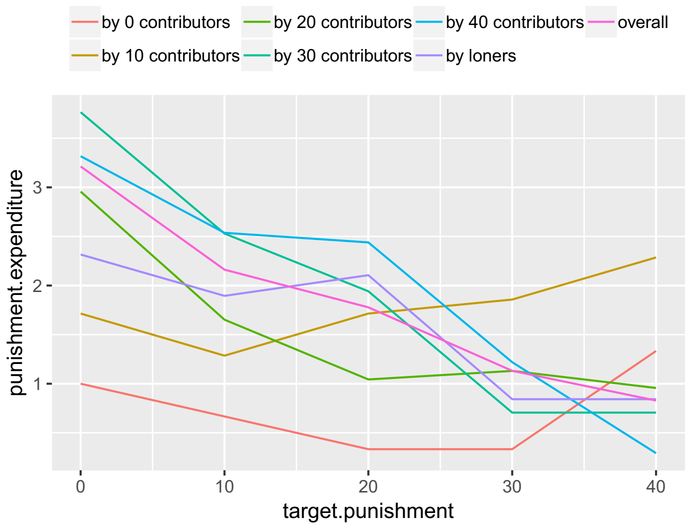


Fig. 5 Compulsory game. Average punishment expenditure by level of contribution directed at all targets, including loners

which increase their punishment level as the target's contribution increases. Similarly to what observed for the optional game, low contributors are responsible for the punishment of high contributors. Moreover, also in this case, loners do not punish high contributors, while contributing only slightly less than the average.

In order to confirm who is responsible for the punishment of high contributors in the compulsory game, we run a regression (Table 5) where the dependent variable is the punishment expenditure of maximal contributors. Findings are, once again similar to those discussed for the optional game: low contributors punish significantly more high contributors, while high contributors do not punish maximal contributors. The results are robust—with and without controls—when we set the punishment expenditure directed at each target contribution and at loners as dependent variable and, as regressors, the contribution level of the punisher (Tables S11–S17 supplementary material). Finally, no significant relationship emerges between loners and the punishment of high contributors.

In conclusion, results provide strong support against HP 5, i.e. that loners punish anti-socially when forced to cooperate. We fail to replicate the results of Rand & Nowak (2011).

4 Discussion

We explored the impact of optional play on contribution and punishment in the public goods game. The combination of these institutional frameworks has been largely overlooked in the experimental literature, which focused mainly on the normative role of punishment as a tool to sustain cooperation. The role of the opt-out strategy in real life relates to all situations in which people can decide not to

Table 5 Punishment of maximal contributors: Compulsory game (logit models)

	Dependent variable		
	40 (maximal)-contributors		
	(1)	(2)	(3)
High contributors (C3)	− 1.287***		
	(0.415)		
Low contributors (C3)		1.431***	
		(0.448)	
Optout(Part.2)			− 0.208
			(0.541)
Age	0.027*	0.028**	0.023*
	(0.014)	(0.014)	(0.013)
Gender	− 0.291	− 0.377	− 0.428
	(0.408)	(0.406)	(0.394)
Trust	− 0.476	− 0.533	− 0.634*
	(0.358)	(0.368)	(0.352)
E1	− 0.215	− 0.203	− 0.201
	(0.454)	(0.448)	(0.438)
Constant	− 0.889	− 1.785*	− 0.991
	(0.962)	(0.995)	(0.928)
Observations	223	223	223
Akaike inf. crit	178.988	179.349	188.874

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

participate rather than free-ride in a common project. The difference between these two behaviours is significant since the person that decides to stay out does not show an intention to take advantage of the others. Relatedly, we also study if and how the introduction of a punishment institution affects the contribution of the community to the common project, with and without the opportunity to opt-out of the interaction.

Previous research (Orbell & Dawes, 1993) suggests that the option to stay out of the game increases cooperation, because it selects people with the real desire to participate—and contribute—to the common project, while those unwilling to contribute stay out of the game. In our experiments, while we see that contributions are higher in the optional game with respect to the compulsory (without punishment), we find this difference being non statistically significant, so we cannot confirm this finding from the literature. Nevertheless, our experiment allows us to replicate the results of Rand & Nowak (2011) that loners contribute significantly less than cooperators when forced to participate.

Concerning punishment, the most important result is that its introduction is very effective to increase cooperation, also with respect to the option to participate. When we introduce this institutional feature significantly more cooperation is observed in the compulsory game, but not in the optional game. We see high cooperation with punishment in the compulsory game, because punishment influences the beliefs of participants about the success of the public goods game and retaliation becomes a credible threat. Participants expect higher contributions because of the presence of punishment and the compulsory nature of participation. Participants trust the punishment institution as a tool to force potential defectors to contribute, so they expect higher contributions and consequently cooperate more on average. At the same time, when punishment is introduced, optional play does not increase cooperation anymore. Optional play suggests that someone can escape the punishment, so participants in the optional game contribute less than in the compulsory game, if there are expectations that loners are not punished as free-riders.

The latter does not seem to be the case: in fact cooperators show similar contempt for the behaviour of loners and defectors, punishing both with the same intensity. At the same time, an interesting picture emerges for loners. When forced to participate, they do not behave as defectors, but their contribution is significantly lower than that of a cooperator in the compulsory game. When we look at their punishment behaviour, we confirm that the ultimate objective of loners is to avoid interactions: they present the lowest possible level of punishment against all contribution levels and, more importantly, they almost never punish other loners, indicating the consciousness of their behaviour. The punishment behavior of loners is consistent with their contribution decision confirming that loners are just subjects who prefer to stay alone, as suggested by Garcia and Traulsen (2012).

We do not find evidence that loners are responsible for anti-social punishment when forced to participate: they do not target high contributors for punishment. Hence, we cannot confirm Rand & Nowak, (2011) result of loners' anti-social punishment.

A speculation about loners emerging from our experiment is that they do not trust others sufficiently to engage in a cooperative enterprise, or they require more personal information about the interacting partner in cooperation (which is excluded in our experimental setup) (Garcia and Traulsen, 2012). The real challenge that emerges from our results is to find the right incentives or social norms that succeed in engaging loners into participation. By observing their punishment behavior, we can exclude that they are anti-social or perverse punishers. Perverse punishment is done chiefly by low contributors. This result is further confirmed because we observe anti-social punishment also when loners are left free to be out of the game. Recent theoretical analysis proposes social norms assigning different social stigma to loners and defectors as vehicles to lead loners to cooperate (Podder et al., 2021).

While our results highlight the importance of studying optional play and punishment together in social dilemma games, further research is needed to analyze these situations in the lab and in the field. Particular attention should be devoted to explore empirically the role of social norms, reputation and information (Podder et al., 2021; Righi & Takács, 2022; Podder et al., 2021) in optional games of cooperation with punishment.

Funding Funding for the experiment was received by Università Degli Studi di Modena e Reggio Emila. SR gratefully acknowledges funding from "Fondi primo insediamento" of the Ca'Foscari University of Venice and from the Italian Minister of University and Research (MUR) under Research Projects of National Relevance (PRIN), project code 2020SKJSTF ("At the frontier of agent-based modelling: a new data driven framework for policy design toward sustainable and resilient economies”). KT acknowledges support of the Hungarian Scientific Research Fund, Grant Number FK 137765688 (PI: Dorottya Kisfalusi).

Data availability statement The dataset generated and analysed in the current study is available from the corresponding author upon request.

Declarations

Conflict of interest Partial financial support was received from University of Modena. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11238-023-09929-9.

References

Aktipis, C. A. (2004). Know when to walk away: Contingent movement and the evolution of cooperation. *Journal of Theoretical Biology, 231*(2), 249–260.

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

Bicchieri, C. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration* (Vol. 3). Princeton University Press.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., & Pancotto, F. (2016). Amoral familism, social capital, or trust? The behavioural foundations of the Italian north-south divide. *The Economic Journal, 126*(594), 1318–1341.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior and Organization, 60*(1), 11–26.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100*(6), 3531–3535.

Brandt, H., Hauert, C., & Sigmund, K. (2006). Punishing and abstaining for public goods. *Proceedings of the National Academy of Sciences, 103*(2), 495–497.

Bravo, G., Squazzoni, F., & Takács, K. (2015). Intermediaries in trust: Indirect reciprocity, incentives, and norms. *Journal of Applied Mathematics, 2015*(S13), 1–12.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2011). Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics, 95*(9–10), 1111–1118.

Castro, L., & Toro, M. A. (2008). Iterated prisoner’s dilemma in an asocial world dominated by loners, not by defectors. *Theoretical Population Biology, 74*(1), 1–5.

Chakravarty, S., & Fonseca, M. A. (2017). Discrimination via exclusion: An experiment on group identity and club goods. *Journal of Public Economic Theory, 19*(1), 244–263.

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

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97.

Chen, X., Szolnoki, A., & Perc, M. (2015). Competition and cooperation among different punishing strategies in the spatial public goods game. *Physical Review E, 92*(1), 012819.

Ehrhart, K.-M., Keser, C., et al. (1999). Mobility and cooperation: On the run. Technical report, Sonderforschungsbereich 504, Universität Mannheim & Sonderforschungsbereich.

Fang, Y., Benko, T. P., Perc, M., Xu, H., & Tan, Q. (2019). Synergistic third-party rewarding and punishment in the public goods game. *Proceedings of the Royal Society A, 475*(2227), 20190349.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

Fowler, J. H. (2005). Altruistic punishment and the origin of cooperation. *Proceedings of the National Academy of Sciences, 102*(19), 7047–7049.

Garcia, J., & Traulsen, A. (2012). Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *Journal of Theoretical Biology, 307*, 168–173.

Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002). Replicator dynamics for optional public good games. *Journal of Theoretical Biology, 218*(2), 187–194.

Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002). Volunteering as red queen mechanism for cooperation in public goods games. *Science, 296*(5570), 1129–1132.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., & Sigmund, K. (2007). Via freedom to coercion: The emergence of costly punishment. *Science, 316*(5833), 1905–1907.

Hauk, E. (2003). Multiple prisoner's dilemma games with (out) an outside option: An experimental study. *Theory and Decision, 54*(3), 207–229.

Helbing, D., Szolnoki, A., Perc, M., & Szabó, G. (2010). Punish, but not too hard: How costly punishment spreads in the spatial public goods game. *New Journal of Physics, 12*(8), 083005.

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science, 319* (5868), 1362–1367.

Liu, L., Chen, X., & Perc, M. (2019). Evolutionary dynamics of cooperation in the public goods game with pool exclusion strategies. *Nonlinear Dynamics, 97*(1), 749–766.

Liu, L., Wang, S., Chen, X., & Perc, M. (2018). Evolutionary dynamics in the public goods games with switching between punishment and exclusion. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 28*(10), 103105.

Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review, 63*(15), 638–660.

My, K. B., & Chalvignac, B. (2010). Voluntary participation and cooperation in a collective-good game. *Journal of Economic Psychology, 31*(4), 705–718.

Nosenzo, D., & Tufano, F. (2017). The effect of voluntary participation on cooperation. *Journal of Economic Behavior and Organization, 142*, 307–319.

Orbell J. M, & Dawes, R. M. (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American Sociological Review*, *58*(6), 787–800. https://doi.org/10.2307/2095951

Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal, 115*(506), 1032–1053.

Pancotto, F., & Righi, S. (2021). Reflectivity relates differently to pro sociality in naïve and strategic subjects. *Scientific Reports, 11*(1), 1–15. https://doi.org/10.1038/s41598-021-91960-3

Phelan, S. E., Arend, R. J., & Seale, D. A. (2005). Using an iterated prisoner's dilemma with exit option to study alliance behavior: Results of a tournament and simulation. *Computational and Mathematical Organization Theory, 11*(4), 339–356.

Podder, S., Righi, S., & Pancotto, F. (2021). Reputation and punishment sustain cooperation in the optional public goods game. *Philosophical Transactions of the Royal Society B, 376*(1838), 20200293. https://doi.org/10.1098/rstb.2020.0293

Podder, S., Righi, S., & Takács, K. (2021). Local reputation, local selection, and the leading eight norms. *Scientific Reports, 11*(1), 1–10. https://doi.org/10.1038/s41598-021-95130-3

Rand, D. G., & Nowak, M. A. (2011). The evolution of antisocial punishment in optional public goods games. *Nature Communications, 2*(1), 1–7.

Righi, S., & Takács, K. (2022) Gossip: Perspective taking to establish cooperation. *Dyn Games Appl*, *12*, 1086–1100. https://doi.org/10.1007/s13235-022-00440-4

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature, 414*(6862), 441–443.

Schuessler, R. (1989). Exit threats and cooperation under anonymity. *Journal of Conflict Resolution, 33*(4), 728–749.

Seale, D. A., Arend, R. J., & Phelan, S. (2006). Modeling alliance activity: Opportunity cost effects and manipulations in an iterated prisoner's dilemma with exit option. *Organizational Behavior and Human Decision Processes, 100*(1), 60–75.

Selten, R., Abbink, K., Buchta, J., & Sadrieh, A. (2003). How to play (3× 3)-games: A strategy method experiment. *Games and Economic Behavior, 45*(1), 19–37.

Semmann, D., Krambeck, H.-J., & Milinski, M. (2003). Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature, 425*(6956), 390–393.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of Biology, 46*(1), 35–57.

Vanberg, V. J., & Congleton, R. D. (1992). Rationality, morality, and exit. *American Political Science Review, 86*(2), 418–431.

Authors and Affiliations

Francesca Pancotto[1] · Simone Righi[2] · Károly Takács[3,4]

✉ Francesca Pancotto
francesca.pancotto@unimore.it

1 Departiment of Culture and Languages, University of Modena and Reggio Emilia, Modena, Italy

2 Department of Economics, Ca' Foscari University of Venice, Venice, Italy

3 Institute for Analytical Sociology, Linköping University, Norrköping, Sweden

4 CSS-RECENS, Centre for Social Sciences, Budapest, Hungary