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

Francesca Pancotto[1], Simone Righi[2,3*], Károly Takács[4,5],

1 Dept. Communication and Economics, University of Modena, Italy

2 Department of Computer Science, University College London, London, UK

3 Ca' Foscari University of Venice, Italy

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

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

* simone.righi@unive.it

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. When opting out of is a feasible alternative, it is questionable whether known solutions to the problem of cooperation, such as punishment could still work, given the limited sanctioning potential it imposes on free riders. We present the results of two experiments with non-student subjects who play optional and compulsory public goods games both with and without a punishment stage. We find that the possibility of opting-out motivates cooperation. Instead, when punishment is introduced, higher cooperation emerges in the compulsory game. This key result indicates that informal solutions to public good problems might rule each other out and punishment is a robust solution only if players are not allowed to opt out of the interaction.

Introduction

The studies of mechanisms and conditions that support cooperation among unrelated individuals are foundational for both social scientists and students of animal behaviour [1, 5]. Social dilemma games, in which individuals face a choice between a cooperative and a non-cooperative action, are among the most important theoretical and experimental tools through which cooperation is studied [6]. However, many real life situations cannot be simplified to a mere choice between cooperating and not cooperating. Individuals often can opt-out from the interaction giving up both potential benefits of cooperation and potential gains from free-riding. For example, a person can decide to join a business alliance [7], to quit a job if he does not like the work environment, to abstain from attending a difficult game of his sports team or the premiere performance of the orchestra. Participation in an international political alliance is also a possible example of cooperation with 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 the one who reaps the benefits of the alliance, without contributing to the balance. Finally, a loner country (the one who decised to stay out of the alliance) is one that instead decides to stay out of it completely, or exit, giving up both the benefits from the alliance and the costs of contributing to the common budget.

Not taking part in a social dilemma interaction has been conceptualized in multiple ways: as exit after play [4, 8, 9]; as walking away [10]; as opting-out [11, 12, 13, 14]; and in a different perspective, when opting out is not voluntary as exclusion [15, 16]. In all these models, 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 [8], the roundrobin elimination of opportunistic counterparts due to their lower overall performance in repeated games [9], and the positive assortment of cooperation strategies favoured by the exit and exclusion mechanisms [10, 15, 16]. Relatedly, when the population of loners is large enough cooperating survives evolutionary pressure [12], leading some authors to suggest that a plausible scenario to start cooperation is a world dominated by loners, not by defectors [17]. 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 [13, 14]. In any case, the relevant parameter for the outcome is the pay-off for refusing to play [18].

Despite the many theoretical arguments proposing that voluntary participation may stimulate cooperation, only a few empirical studies have so far examined the effects of voluntary participation on cooperation [11, 19, 20, 21, 22]. These studies compare the effects of entry and exit options on cooperation, using public goods game experiment and showing that the exit option is capable of sustaining cooperation through the value of the threat (while the entry strategy is not effective [19]). These results highlight that exit and opting-out are efficient institutions for enhancing cooperation when they can be considered as credible threats, i.e. there are other interactions or benefits outside the current interactions. There is no convincing evidence of how could opting-out enhance cooperation in the absence of outside options.

In this study, we highlight that the availability of another institution, punishment, sheds light on the motivations behind the choice to opt-out. This differentiates our contribution from experimental exercises in which there is a voluntary choice in the selection of a contribution out of different options [23] and from endogenous group formation [24, 41]. In fact, none of the previous contributions has explored the interaction between opting-out and punishment behaviour, which is what we do.

The consideration of this interaction is not only important for finding out the true motivational background behind the choice to opt-out, but also to gain a deeper understanding of the conditions under which punishment and opt-out could be efficient institutions for solving social dilemmas. Post-interaction punishment alone has been demonstrated to provide a powerful solution, both by theoretical [27, 28, 29] and experimental work [30, 31]. It remains an open question, however, if opting-out could further improve the efficiency of the punishment institution. When punishment is interacted with opting-out, first one needs to consider that the set of possible punishment strategies is increased. Predictions might differ if only certain punishment strategies are allowed, for instance, when players can punish defectors but not cooperators [32]. The success of cooperators who also take the cost of punishing defectors depends on the presence of the opt-out strategy that provides larger payoffs than the the payoff of defectors who are punished [33]). If models are extended to the possibility allowing the punishment of all strategies, the dynamics leads to cyclic behaviours that do not allow any prediction concerning the sustainment of cooperation

[25, 34]. This cyclic behaviour has been confirmed also in the laboratory [35].

In the present paper, we study how the introduction of punishment impacts on 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 option to participate and we run them with a sample of the general population. A similar design to ours is the one devised by Rand and Nowak [25], where they test 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 [26]. Similarly to us, [19] 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 no-play or opt-out option, the interaction between these two elements is an important path to follow. It is important to study the two institutions together, because whether punishment is effective 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 see who increases or decreases his/her contributions in the different institutional settings.

We design our experiments to test the following hypotheses:

Hypothesis **1** *Optional play induces higher cooperation in a public goods game without punishment.*

Hypothesis **2** *Compulsory play induces higher cooperation in a public goods game with punishment.*

Hypothesis **3** *Loners are low contributors when forced to participate in a public goods game (HP1 in [25]).*

Hypothesis **4** *Loners are anti-social punishers when forced to participate in a public goods game (HP2 in [25]).*

To test our hypotheses, 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 [25], allows to identify those who decide to not participate and compares 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 the 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 accept 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 [25] hypothesis on the behaviour of loners, with a sample taken from the general population.

Materials and methods

Experimental design

In order to test our hypotheses, we conduct two experiments, E1 and E2, on different subjects (Total N = 236). The exact composition of the sample is reported in Table S18 of Supplementary Information (SI). As the problem of cooperation is intrinsically linked with the problem of social norms [2, 36] and involve society as a whole, we select a sample from the general national population at the time of the experiment (recruitment procedures are described in the *Methods* section). However due to limitations in the pool of available potential participants, the actual sample overrepresents 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 discrepancy with respect to the underlying population. Despite these limitations, a broad spectrum of the population is covered by the sample, unlike in typical lab experiments with university students.

	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%

Table 1. Sample composition discrepancy with respect to national average

The public goods game We use a discrete form of PGG game in both experiments. All group members receive an endowment of $e = 40$ experimental points and have to decide simultaneously how much of their endowment to invest in a common project, choosing a contribution level among the following possible values, ($c_i \in 0, 10, 20, 30, 40$), knowing that the residual ($e - c_i$) would remain in their private account. Every point invested in the common project is then doubled and shared equally among group members. Hence, individual earnings are 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. Differently from the case of a standard PGG, here participants can also stay out of the game: by staying out, their reward is their initial endowment $c_i = 40$, while earnings from the public good are divided, in equal parts, only among participants. Consequently, the MPCR decreases as a function of the number of participants, given that the multiplication factor remains at 2 also when some opt-out.

For $N = 0$, nobody participates and the PGG is not put in place. Notably, the case of nobody participating to the PGG and the case in which all participants contribute

Label	Part	Decision	PGG-Game
Part 1:	First Part	Participation	Optional
C1.1 :	First Part	Contribution	
C1.2:	First Part	Contribution	Compulsory
Part 2:	Second Part:	Participation	Optional
C2 :	Second Part:	Contribution	
PUN2:	Second Part:	Punishment	Optional
C3:	Third Part:	Contribution	Compulsory
PUN3:	Third Part:	Punishment	Optional

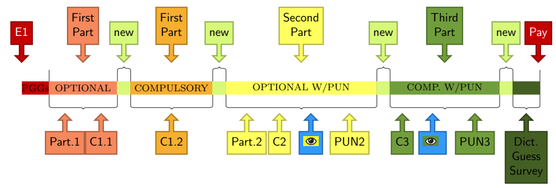
Table 2. Legend of the time line

$c_i = 0$ are *ex-post* payoff equivalent for a group member. Indeed, in both cases both loners and free-riders (zero contributors) obtain the same payoff, $c_i = 40$. Before knowing how many will contribute to the PGG, however, free-riders may expect a positive return from the PGG, while the loners' payoff is fixed and independent from the decision of others. This consideration highlights how loners might lack trust towards other group members: from an individual payoff point of view, it would be always more convenient to participate to the public goods game and free ride, rather than to stay out of it entirely.

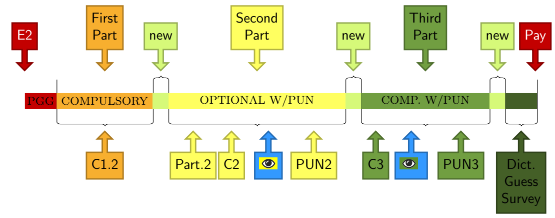
The discrete contributions framework is chosen with the aim to facilitate calculations and understanding of the game for the participants. The time line of experiments is reported respectively in Fig. 1a for E1 and Fig. 1b for E2. The legend of both is in Table 2.

Experiment 1 The first experiment (E1), involving 176 participants, is composed of three parts. In the First part, an Optional Public Goods Game (O-PGG) is followed by a Compulsory Public Goods Game(C-PGG), with PGG parameters as described. Between the optional and the compulsory games, groups are reshuffled and participants are informed that they will play with new – different – group members. Indeed, in order to obtain a stranger matching protocol, every time the groups play a different PGG groups are reshuffled (in Figs. 1a and 1b, the 'new' light-green sign indicates that there is a group reshuffling).

In the Optional game, after explaining the rules of the PGG, subjects can decide to exit the game, in exchange for a fixed amount of points, $c_i = 40$. This payoff corresponds to the loners' payoff in [25] model, where a loner gets a payoff which is



(a) Experiment 1 (E1) timeline



(b) Experiment 2 (E2) timeline

Fig 1. *Note: The eye symbol indicates the observation of the decisions of group members*

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 (Part 1 in Fig. 1a). Then, subjects who decided to participate, decide their contribution level (labelled C1.1). After reshuffling, subjects are asked to make a compulsory contributions 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 C1.2.

After re-matching, we run the second part of the experiment, consisting of the Optional PGG with Punishment (O-PGG-W/P). The O-PGG-W/P, is composed of three decisions: first subjects have to decide whether to participate in the game (Part 2 in the time line). Only those who decide to participate make a contribution decision for the PGG (C2). After C2, all subjects in the group observe the decisions of other group members: this is indicated with the eye symbol in the time line. This is the only part where subjects receive a feedback related to the experiment, except the very end. In the third decision of the second part, all subjects have a punishment stage, in which everyone has to express whether and how much they want to punish each possible contribution level, as well as the exit option, at a cost for themselves. This decision is indicated as PUN2 in Fig. 1a. Participants can select one of three punishment levels: no punishment at zero cost, reducing the payoff of another participant with 12 points at a cost of 4 for the punisher, and reducing the payoff of another participant with 24 at a cost of 8 for the punisher. This punishment structure correspond to a 1:3 punishment technology.

We stress that participants are allowed to punish also those who opt-out (loners). Loners are also asked to make a punishment decision toward other group members, which metans that we move beyond the [25] experimental design. The punishment decision that counts for the final payment is selected randomly by the computer, which means that everyone gives his potential punishment decision for each level of potential contribution of other members, but then only one choice is implemented, through a random matching with another group member.

It is important to note that in the punishment stage, each group member selects a punishment level for each possible contribution level, rather than selecting a punishment level for specific group members. This punishment method allows to obtain information on behavioral response to all possible decisions of others and help to identify the strategies participants follow. In our design, we will also be able to assess the responses of loners, who are similar to third party punishers [3, 37] that observe the society and judge behaviours without being directly involved in it, however differently to third party punishers, the loners can be themselves be punished. It is worth noticing that, thanks to this punishment structure, subjects are always allowed to decide not to punish, so the punishment decision is always an optional decision. Finally, in the third part of E1 we run a Compulsory PGG with Punishment (C-PGG-W/P) which is identical to the game of the second part, but without the possibility to stay out of the game (C3). Participants observe the contribution levels of group members and then decide 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.

Experiment 2 In the second experiment (E2), run with 60 subjects different from the previous ones, we remove the optional game followed by a compulsory game in the first part and substitute it with a simple PGG without the possibility of opt-out. The remaining parts of the design are 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 Figs. 1a and 1b).

Results

Analysis of Contributions

We first report overall contribution rates, together with a Mann-Whitney test (M-W) for the significance of differences in means between the contributions, calculated by experimental stages and overall. Results of the two experiments are comparable as the difference between overall average contributions is not significant (E1= 26.45 vs E2= 25.97, M-W, p-value(p)=0.189, *one-sided*; Table 3, first line). We also compare contributions in the different parts of the game between E1 and E2: the only significant difference is in the one observed in the Optional Game with Punishment (C2), which is significantly larger in E1 than in E2 (M-W, p=0.041). The average contributions (Table 3, first column) change in the different parts of the game, as a result of the change in the institution presented to the experimental subjects: in the game without punishment, eliminating the option to stay out of the game decreases average contributions (from C1.1=25.91 to C1.2=25.23), but this difference is not statistically significant (Table 4, M-W p=0.36). While the direction of the change goes in the direction as suggested by hypothesis 1, the observed difference is not statistically significant. Hypothesis 2 is instead supported given that in the game with punishment the elimination of the opt-out option increases contributions in the compulsory game (from C2=26.41 to C3=27.62; M-W p=0.097; Tables 3 column 1, final rows and 4 last column).

Avg. Contributions	E1+E2 N=236	E1 N=176	M-W E1 vs E2	E2 N=60
Overall	26.35	26.45	p=0.18	25.97
C1.1	25.91	25.91	-	-
C1.2	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

Table 3. Comparison of Contributions levels across parts of the game

We can thus conclude that:

EVIDENCE **1** *Introducing compulsory play increases cooperation in the public goods*

E1+E2	C1.1(=25.91)	C1.2(=25.23)	C2(=26.41)	C3(=27.62)
C1.1	-	0.36	0.274	0.031
C1.2	-	-	0.173	0.001
C2	-	-	-	0.097
C3	-	-	-	-
		C2+C3 (=27.07)		
C1.1 + C1.2(=25.75)		0.016		

Table 4. Average contributions and Mann Whitney Test (p-values) of the significant difference in average contributions between stages.

game with punishment.

More in general, we observe in Table 4 that the average contribution (25.75) is significantly higher in all the games with punishment with respect to the games without punishment (27.07, M-W, p=0.016). Moreover, the highest contribution level is observed when both compulsory contribution and punishment are present. Indeed, in Table 4 C1.1 (=25.91) is smaller than C3 (=27.62) and this difference is significant (M-W, p=0.031). In the second row, we find that C1.2(=25.23) is significantly smaller than C3 (M-W; p=0.001) and in the third row that the average contribution in C2(=26.41) is significantly smaller than C3 (M-W; p=0.097). Unexpectedly, we find that the punishment institution enhances its effectiveness in eliciting cooperation when introduced in a compulsory play setup. To sum up:

Evidence **2** *Contributions are significantly larger in games with punishment,ì than in games without. Contributions are highest when both punishment and compulsory contributions are present.*

Comparison between the contribution levels of participants and loners in compulsory games. In this Section, we explore further the result summarized in Evidence 2, where we found support for Hypothesis 2 and we showed that the highest level of cooperation are found in the compulsory game with punishment.

To this end, we separate average contribution levels between Participants and Loners (Table 5), and look at how these two groups respond to different institutional structures. Loners contribute significantly less than Participants in the compulsory game both without punishment (C1.2: Participants= 25.73, Loners= 20.83, M-W, p=0.09) and with punishment (C3: Participants=28.33, Loners=23.94 , M-W, p=0.02). These results hold when looking at E1 and E2 separately (See Table Table 5), so that we can state that:

EVIDENCE **3** *Loners contribute significantly less than Participants in a compulsory game, showing a different willingness to cooperate.*

This result supports our HP 3 and confirms the results of [25]. When Loners are forced to participate, their contribution is larger than zero, so they cannot be assimilated to free-riders. Furthermore, when punishment is introduced, loners increase slightly their contribution, but in a non-significant way (C1.2=20.83→C3=24.69, p=0.14). It is clear that loners are not strongly affected by the introduction of punishment (Loners: C1.2=20.83 to C3=23.94, M-W p=0.16). Figure S1 confirms both results visually. Given Evidence 3, we can observe that the aggregate contributions

E1+E2	All subjects	N. part	Participants	M-W p	Loners	N loners
C1.1	25.91	164	25.91		-	12
			(↓ p=0.56)			
C1.2	25.33	-	25.73	(→p=0.090)	20.83	
C2	26.41	198	26.41		(↓ p=0.16)	38
	-		(↓ p=0.03)			
C3	27.63	-	28.33	(→p=0.02)	23.94	
E1	All subjects	N Part.	Participants	M-W p	Loners	N loners
C1.1	25.91	164	25.91	-	-	12
			(↓ p=0.56)			
C1.2	25.41		25.73	(→p=0.090)	20.83	-
C2	27.15	144	27.15	-	(↓ p=0.14)	32
			(↓ p=0.21)			
C3	27.44		28.06	(→p=0.098)	24.69	-
E2	All subjects	N Part.	Participants	M-W p	Loners	N loners
C1.2	25.16		25.74	(→p=0.08)	20	-
C2	24.44	54	24.44	-	-	6
			(↓ p=0.009)	-	-	
C3	28.16		29.07	(→p=0.037)	20	-

Table 5. Average contributions of participants and loners

observed are strongly influenced by contribution levels of participants. Comparing the optional (C11=25.91) and compulsory games (C12=25.73) without punishment, the total contribution decreases as an effect of the reduction of participants' contributions.

In the game with punishment, participants increase consistently their contributions when a compulsory play is introduced (C2=26.41 to C3=28.33, M-W p=0.033). 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, induced by the punishment institution. This is reflected by survey response to a question on trust (formulated as in the World Value Survey), which becomes a significant in explaining contribution in compulsory play when punishment is present (Tables S1 and S2 in SI). Participants contribute significantly more and show higher level of trust in the game with punishment, but not in the game without punishment. To study the evolution of behaviours of different types of agents, we attribute artificially the value of 0 to the contribution of Loners in the optional games and calculate the variation (delta) between the contribution levels in the two games as the difference between compulsory game and the relative optional game (Fig.2, from data in Table S1). Thus, a positive variation indicates that the subject has increased his/her contribution when passing from the compulsory to the optional game. On the one side, participants show more often a positive delta and a lower level of constant contribution in the game with punishment than they do in the game without. On the other side, the distributions of the loners' deltas are quite similar in the two games, as suggested by the result of the MW tests and confirming that they are rather unaffected by the punishment institution.

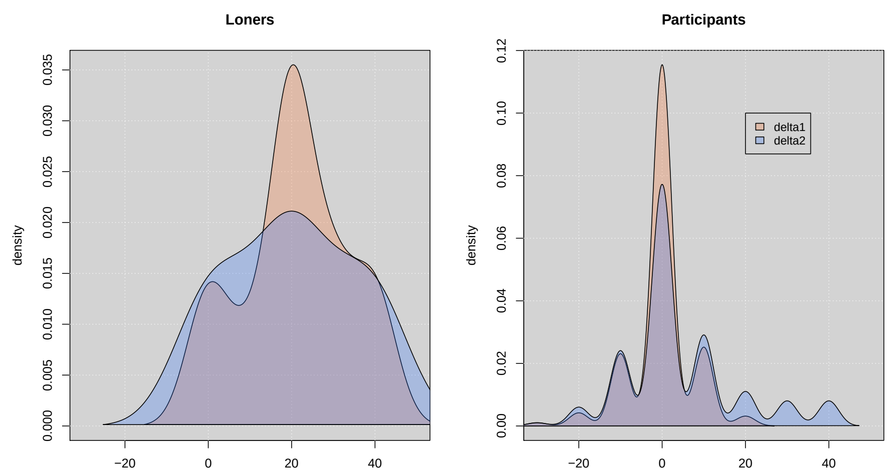


Fig 2. Distributions of delta contributions in both games. A positive delta indicates an increase of contribution from the Optional to the Compulsory PGG. delta1 relates to the PGG without Punishment, delta2 to the PGG with Punishment.

Punishment

We now report the characteristics of punishment behavior in the optional and in the compulsory games in more detail. We explore also the presence of anti-social punishment, defined by [26] as the punishment in which the punisher contributes less than, or the same amount, as the target of his/her punishment. A second analogous definition [38] distinguishes between 'normal' and 'perverse' punishment, where 'perverse' punishment is the one directed at the highest contributors (or at someone who has contributed more than the average): in our case the punishment toward contribution levels equal to 30 and 40 is *perverse* and the punishment directed at contribution levels equal to 0 or 10 is *normal*. Further, we study punishment of maximal contributors, i.e. those contributing exactly 40.

Optional game In the optional game, the average punishment expenditure invested by agents expressing each level of contribution and directed at all contribution levels and at loners is reported in Fig. 3 (numerical data in Tables S4 and S5). Overall punishment expenditure is decreasing with the increase in the target's contribution level. 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 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. Notably, loners are allowed to stay out of the game, thus they cannot be be responsible for it. 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. We conclude that:

EVIDENCE **4** *We observe perverse (anti-social) punishment in the optional public goods game, where loners are not forced to participate in the contribution. Low contributors engage in perverse punishment.*

The punishment pattern is strikingly different for loners, who engage in very low levels of punishment toward maximal contributors (violet line in Figure 3, precise values

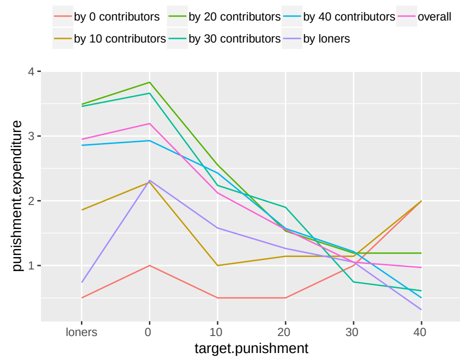


Fig 3. Optional Game. Average punishment expenditure by level of contribution directed at all targets, including loners.

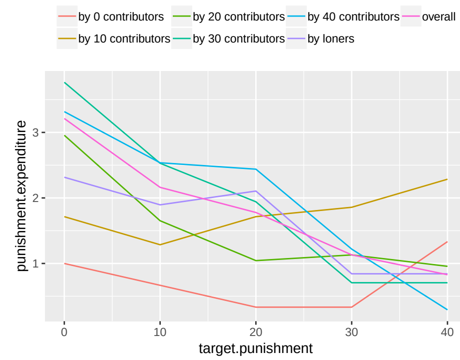


Fig 4. ompulsory game. Average punishment expenditure by level of contribution directed at all targets, including loners.

in Table S4). Moreover the punishment of loners by all contribution levels (Overall line), is affine to the one of low contributors. The behaviour of loners is associated by the group to the one of low contributors, and this effect is stronger for higher contributors.

The punishment of loners by loners is also low, helping to shed light on the opinion that this type of agent has about its own strategy. Indeed, loners do not punish other loners i.e. they support their decision by abstaining from punishing behaviours similar to their own. Finally, loners punish on average less than the other participants, confirming that their main objective is to avoid interactions with their group. Summarizing evidence concerning loners, we can state that:

Evidence **5** *In the optional game, loners do not punish high contributors. Loners do not punish other loners. Loners are punished similarly to low contributors.*

Studying the results of multivariate logit models on the decision to punish maximal contributors and loners controlling for age, gender, and trust (Table 6), w e 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 the contribution level of the punisher; both with and without controls (Tables S6-S11). Further, from Table 6, we confirm that the decision to opt-out is significantly and negatively related to the punishment of maximal contributors, confirming that loners are not responsible for perverse punishment (Evidence 4).

Table 6. 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
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 7. 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
Note:	*p<0.1; **p<0.05; ***p<0.01		

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 5).

Compulsory game Analyzing the average punishment expenditure by the level of contribution in the compulsory game (Figure 4), we observe a decreasing trend of punishment as a function of the contribution of the target. For high and medium contributors, the higher the contribution of the punisher, the lower the punishment of high contributors, i.e., punishment is on average normal. However, the opposite happens for low contributors, which increase their punishment level as the contribution of the target increases. Just like in the optional game, low contributors are responsible for the punishment of high contributors. Moreover, again loners do not punish high contributors, they behave like average contributors, although contributing slightly less than the average.

In order to confirm who is responsible of punishment of high contributors in the compulsory game, we run a regression (Table 7) with as dependent variable the punishment expenditure of maximal contributors. We find similar results to those discussed for the optional game: low contributors punish significantly more high contributors, while high contributors do not punish maximal contributors. Again, the results are robust when we adopt as dependent variable the punishment expenditure directed at each target contribution and at loners and as regressors the contribution level of the punisher; both with and with and without controls (Tables S11-S17). Finally, no significant relationship emerge between loners and the punishment of high contributors .

Results provide strong support against our HP 4, i.e. that loners punish anti-socially when forced to cooperate. We fail to replicate the results of [25].

Discussion

We explored the impact of an opt-out option on contribution and punishment in a public goods game. The importance of this institutional framework has been largely overlooked in the experimental literature, which has 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 those situations in which people can decide to not participate rather than defect in a common project: the person that decides to stay out does not have the intention to take advantage of the others, but rather to express the lack of confidence in the reference community. Relatedly, we also study if and how the introduction of a punishment institution, affects the contribution of the community to the common project.

We find that the option to stay out of the game increases cooperation in the game without punishment. [11] suggest an explanation for this evidence: allowing subjects to opt-out of a common project enhances cooperative behaviour, because it effectively selects people with the real desire to participate - and contribute - to the common project, while those unwilling to participate stay out of the game. Those who participate in the optional game, are induced to contribute more, expecting a higher likelihood of meeting mainly peers who are really willing to participate. Moreover, we replicate the result of [25], because in our data loners contribute less than other players 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 punishment is introduced, more cooperation is observed in the compulsory game, but not in the optional game. We see high cooperation with punishment in the mandatory game, because punishment influences the beliefs of participants about the success of the public goods game and retaliation is a credible threat. Participants expect higher social contributions because of the presence of punishment and the compulsory nature of the contribution. 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.

Participants show their opinion on the behaviour of loners in the punishment section, punishing them with the same intensity as they punish defectors. 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 participant in the compulsory game. When we look at their punishment behaviour, we confirm the willingness of loners to avoid interactions: they present the lowest possible level of punishment against all possible contribution levels and, more importantly, they significantly punish less other loners, indicating the consciousness of their behaviour. We do not find evidence that loners are responsible of anti-social punishment when forced to participate: they do not target high contributors for punishment. Hence, we cannot confirm [25] result of loners' anti-social punishment: in their model the authors indicate that a form of information leakage, thanks to which loners are informed of the identity of invading cooperators, is needed for their anti-social behaviour. In our experimental setting, this information leakage is excluded and this could be a reason why we do not observe the same results. The punishment behavior of loners is consistent with their contribution decision: that is loners are just subjects who prefer to stay alone, as suggested by [39]. Loners might be different people who do not trust enough their society to engage in a cooperative enterprise, or they require personal knowledge for cooperating (which is excluded in our experimental setup)[39]. The real challenge that emerges from our results is to find the right incentive or social norm that succeeds in engaging loners into the society. Indeed, by observing their punishment behavior, we can exclude that they are anti-social or perverse punishers. The latter type of punishment is done chiefly by low contributors. And this result is further confirmed because we observe anti-social punishment also when loners are left free to be out of the game.

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 and to explore the role of reputation and information in optional games of cooperation.

Supporting information

Methods

The experiment was conducted with the use of a mobile laboratory for experimental economics. 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 province), Macerata (in the main town 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. The sessions in Macerata were run at the Department of Economics of the University; the sessions in Fiastra were run on a container made available by a local association. All venues used for the experiment were all accessible by car or public transport.

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 municipality newsletter and a Facebook page. 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.

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 municipalities in which the session was run, or of the neighbouring ones, could participate in an experiment run there (locations of the sessions, dates, times are reported in Table S19).

The experimental software was developed in Python using the o-Tree platform [40]. This software platform was chosen as it allows 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).

Upon arrival at the experimental session, subjects were registered and assigned a seat where they were given both an informed consent as well as a privacy consent and data release form to read and sign. Participants were informed 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) experimenter. Moreover, the instructions for the current game were available at the bottom of the screen at any time during the associated task.

The experiment was 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 therefore approved by the REBEL' ethics committee.

Average session time was one hour, and payoffs were expressed in experimental points (tokens), with each token corresponding to 0.04 €. The average payments per person was around 15 Euros. The composition of our final sample as well as general characteristics of the towns selected are described respectively in Table S18 and S19 of SI. The number of participants was 16 for each experimental session, with every individual allowed to participate in only one session. In some cases due to last minute abandonment, we run sessions of 12 subjects. All sessions were run on Saturdays in order to favour a wider and more diverse participation.

Sample selection For Experiment (E1), 176 subjects were recruited according to the procedures explained in the previous section. We did not proceed to exclude subjects that did not respond correctly to the answers, given the complex nature of the recruiting procedure, but we did give feedback of the control questions to the participants. When the answer was given, in the next page of the software, a calculated explanation was provided in order to help the participants to fully understand the game structure. Moreover, we collected the data related to the performance of the participants in the game and used the variable as a control in the analysis. Subjects received a 5 Euros show up fee and on average extra 15 Euros for the decisions made in the experiment. In each part of the experiment, new groups were formed with a reshuffle among all participants in the session. Participants were informed of the reshuffling at all stages in which it occurred. First the experimenter read a set of instructions for a one shot public goods game in which the general rule of functioning of a public goods game is explained. Subjects were then informed that they entered the second part of the experiment, with new group members. Instructions were read aloud by the experimenter that explained that the game was similar to the previous part but that there would be a second phase of the game where subjects should choose among a list of possible decisions to make. In each part, when a new game is introduced, subjects were asked to respond to 3-4 control questions, which would not affect the game earnings. In order to prevent between-game learning, we did not provide any feedback related to the games before the end of the whole experiment. The only part where a feedback is provided, is in the PGG games with punishment, where the contribution levels and participations are shown before making a punishment decision. In the second experiment, E2, 60 subjects were recruited. The only difference with respect to E1, is in the first part of the game. All the rest is identical, as shown in Figures 1b and 1b. 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 last games are not analysed in this paper. 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 points in Euros. Payments were made privately and in cash at the end of the session.

Details on Punishment The Punishment stage of our design is the same for both E1 and E2. 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.

Subjects are allowed to condition their Stage 2 choice on the other player's contribution in the previous stage (the PGG contribution decision): although they can observe the amount contributed by the other players in the group, they cannot not directly target their punishment at the subjects with a specific level of contribution. On the contrary, subject 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 the choice to not participate to the game. 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. Subjects were informed of this procedure and payment structure all along. The translation of the experimental instructions are provided in Section S4 of SI.

Data availability statement

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

Acknowledgments

References

1. Axelrod R. The complexity of cooperation: Agent-based models of competition and collaboration. vol. 3. Princeton University Press; 1997.

2. The grammar of society: The nature and dynamics of social norms Bicchieri, Cristina 2005,Cambridge University Press

3. Intermediaries in trust: Indirect reciprocity, incentives, and norms Bravo, Giangiacomo and Squazzoni, Flaminio and Takács, Károly Journal of Applied Mathematics 2015

4. Rationality, morality, and exit Vanberg, Viktor J and Congleton, Roger D American Political Science Review, 86 (2), 418-431, 1991 Cambridge University Press

5. Riolo RL, Cohen MD, Axelrod R. Evolution of cooperation without reciprocity. Nature. 2001;414(6862):441.

6. Trivers RL. The evolution of reciprocal altruism. The Quarterly review of biology. 1971;46(1):35–57.

7. Seale DA, Arend RJ, Phelan S. Modeling alliance activity: Opportunity cost effects and manipulations in an iterated prisoner's dilemma with exit option. Organizational Behavior and Human Decision Processes. 2006;100(1):60–75.

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

9. Phelan SE, Arend RJ, Seale DA. Using an iterated Prisoner's Dilemma with exit option to study alliance behavior: Results of a tournament and simulation. Computational & Mathematical Organization Theory. 2005;11(4):339–356.

10. Aktipis CA. Know when to walk away: contingent movement and the evolution of cooperation. Journal of theoretical biology. 2004;231(2):249–260.

11. Orbell JM, Dawes RM. Social welfare, cooperators' advantage, and the option of not playing the game. American sociological review. 1993;58(6):787–800.

12. Batali J, Kitcher P. Evolution of altriusm in optional and compulsory games. Journal of theoretical biology. 1995;175(2):161–171.

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

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

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

16. Liu L, Chen X, Perc M. Evolutionary dynamics of cooperation in the public goods game with pool exclusion strategies. Nonlinear Dynamics. 2019; p. 1–18.

17. Castro L, Toro MA. Iterated prisoner's dilemma in an asocial world dominated by loners, not by defectors. Theoretical population biology. 2008;74(1):1–5.

18. Macy MW, Skvoretz J. The evolution of trust and cooperation between strangers: A computational model. American Sociological Review. 1998; p. 638–660.

19. Nosenzo D, Tufano F. The effect of voluntary participation on cooperation. Journal of Economic Behavior & Organization. 2017;142:307–319.

20. Ehrhart KM, Keser C, et al. Mobility and cooperation: On the run. Sonderforschungsbereich 504, Universität Mannheim & Sonderforschungsbereich . . . ; 1999.

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

22. My KB, Chalvignac B. Voluntary participation and cooperation in a collective-good game. Journal of Economic Psychology. 2010;31(4):705–718.

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

24. Brekke KA, Hauge KE, Lind JT, Nyborg K. Playing with the good guys. A public good game with endogenous group formation. Journal of Public Economics. 2011;95(9-10):1111–1118.

25. Rand DG, Nowak MA. The evolution of antisocial punishment in optional public goods games. Nature communications. 2011;2:434.

26. Herrmann B, Thöni C, Gächter S. Antisocial punishment across societies. Science. 2008;319(5868):1362–1367.

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

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

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

30. Fehr E, Gachter S. Cooperation and punishment in public goods experiments. American Economic Review. 2000;90(4):980–994.

31. Boyd R, Gintis H, Bowles S, Richerson PJ. The evolution of altruistic punishment. Proceedings of the National Academy of Sciences. 2003;100(6):3531–3535.

32. Fowler JH. Altruistic punishment and the origin of cooperation. Proceedings of the National Academy of Sciences. 2005;102(19):7047–7049.

33. Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K. Via freedom to coercion: the emergence of costly punishment. science. 2007;316(5833):1905–1907.

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

35. Semmann D, Krambeck HJ, Milinski M. Volunteering leads to rock–paper–scissors dynamics in a public goods game. Nature. 2003;425(6956):390.

36. Bigoni M, Bortolotti S, Casari M, Gambetta D, Pancotto F. Amoral Familism, Social Capital, or Trust? The Behavioural Foundations of the Italian

North–South Divide. The Economic Journal. 2016;126(594):1318–1341. doi:10.1111/ecoj.12292.

37. Fehr E, Fischbacher U. Third-party punishment and social norms. Evolution and human behavior. 2004;25(2):63–87.

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

39. Garcia J, Traulsen A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. Journal of theoretical biology. 2012;307:168–173.

40. Chen DL, Schonger M, Wickens C. oTree. An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance. 2016;9:88 – 97. doi:http://dx.doi.org/10.1016/j.jbef.2015.12.001.

41. Page T, Putterman L, Unel B. Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. The Economic Journal. 2005;506:1032–1053.