Coordinated and uncoordinated punishment in a team investment game

Vicente Calabuig[1] · Natalia Jiménez-Jiménez[2] · Gonzalo Olcina[1] ·
Ismael Rodriguez-Lara[3,4]

Abstract

Coordinated punishment occurs when punishment requires a specific number of punishers to be effective, otherwise, no damage will be inflicted on the target. While societies often rely on this punishment device, its benefits are unclear compared to uncoordinated punishment, where punishment decisions are substitutes. In this paper, we compare the efficacy of coordinated and uncoordinated punishment in a team investment game with two investors and one allocator. Our findings indicate that coordinated punishment results in higher levels of cooperation and reciprocity, as measured by the levels of joint investment and the return by allocators. Importantly, this does not translate into higher payoffs: investors use punishment more frequently when this is coordinated, which destroys the efficiency gains generated by the highest investment. In fact, our results suggest that the highest level of efficiency would be achieved if investors were not allowed to punish.

A previous version of this paper has circulated under the title "United we stand: On the benefits of coordinated punishment".

✉ Ismael Rodriguez-Lara
  ismaelrl@uma.es

1   ERI-CES and Department of Economic Analysis, Facultad de Economia, University of Valencia, Campus dels Tarongers, Avda. dels Tarongers s/n, 46022 Valencia, Spain

2   Departamento de Economía, Métodos Cuantitativos de Historia Económica, Universidad Pablo de Olavide, Edificio 3, 3ª Planta, despacho 19 Carretera de Utrera, km 1, 41013 Seville, Spain

3   Departamento de Teoría e Historia Económica, Facultad de Ciencias Económicas, Universidad de Malaga, Campus El Ejido, 29013 Málaga, Spain

4   Economic Science Institute (ESI), Argyros School of Business and Economics, Chapman University, One University Drive, Orange, CA 92866, USA

⌬ Springer

1 Introduction

To be successful, individuals may need to coordinate their actions. For example, consider the case of two employees who decide to engage in strike action to negotiate pay and condition agreements with their employer. Although there are other instances of what we refer to as *coordinated punishment* (e.g., think of partnerships, professional societies, or unions), most existing models and experimental research typically assume that punishment is uncoordinated. In particular, it is often assumed that punishment decisions are substitutes and punishment is carried out on an individual basis. Arguably, this way of modelling punishment leaves a plethora of situations unexplained in which members of a group cannot, by themselves, inflict any damage. There are indeed cases in which punishment will not occur unless there is unanimity, e.g., most States of the United States juries require unanimity for finding a defendant guilty.

This paper compares the effectiveness of coordinated and uncoordinated punishment in a team investment game. The focus is on asymmetric social interaction based on division of labour or agency settings, e.g., our game can be thought of as a hold-up game under incomplete contracts with several investors. We opted for a team investment where the proceeds of the aggregate investment of the team members (investors) are under the control of another agent (the allocator), because such situations are quite ubiquitous in real economies. The labour market provides a prominent example. In numerous employment arrangements, a group of workers are employed by a single employer (the firm). The employment agreement in such scenarios is highly incomplete and tends to confer substantial authority on the employer (Baron & Kreps, 1999). This asymmetric distribution of decision rights puts employees at risk of exploitation and leads to inefficiencies if they refuse to cooperate (Gambetta, 2000). In cases of team investment, investors (workers) face the collective action problem of credibly threatening to punish the allocator's (the firm's) opportunistic behavior.

We consider a team investment game with two investors and one allocator. The investors are considered to be workers in a firm or partners in a joint venture who decide whether or not to invest effort in a joint project, while the allocator decides how to share the investment's returns. We assume that investment decisions are complements (e.g., Brandts & Cooper, 2006; Harrison & Hirshleifer, 1989; Riedl et al., 2015; Van Huyck et al., 1990) and that the allocator values the investors' decisions equally. This implies that the allocator will return the same amount to them if there is joint investment. Investors decide whether or not to punish the allocator based on the amount returned.

In our design (between-subjects), we consider four different treatments, depending on the punishment device.[1] When punishment is uncoordinated, the allocator's payoffs are reduced by 30% (60%) if one (both) of the investors decides (decide) to

---

[1] The fact that we impose complementarity in the investment decisions and split the allocators' return into two identical parts implies that both investors are in the same position when deciding whether or not to punish.

punish. Punishment requires coordination to be effective in the case of coordinated punishment. In particular, the allocator's payoffs in this treatment are not reduced if only one investor punishes, while the allocator's payoffs are reduced by 60% if both investors punish.[2] We note that one possible reason for a better performance of coordinated punishment, according to some authors (Boyd et al., 2010; Olcina & Calabuig, 2015) is that coordinated punishment is associated with *increasing returns to scale.* We incorporate this feature in a treatment where the allocator's payoffs are not reduced if only one investor punishes, while the allocator's payoffs are reduced by 80% if both investors punish. Finally, we consider a treatment in which punishment is not possible, as outlined in Cassar and Rigdon (2011). This treatment is used as a benchmark scenario to explore the impacts of punishment on behavior in a team investment game, which has not yet been investigated in the literature.

We provide experimental evidence that the levels of joint investment and the allocator's return are higher when punishment has to be coordinated, compared with uncoordinated punishment. We also find that increasing returns to scale in coordinated punishment helps in fostering the level of joint investment but does not seem to influence the allocators' return. This, in turn, implies that coordinated punishment might perform better than uncoordinated punishment, even if returns to scale are absent. One might expect these findings to translate into efficiency gains when punishment has to be coordinated. Empirically, this does not appear to be the case. If we look at the sum of total payoffs, there are no differences across treatments when we compare coordinated and uncoordinated punishment. One important aspect that contributes to this finding is the fact that investors rely more frequently on joint punishment when punishment has to be coordinated, i.e., before punishing, total payoffs are higher when punishment is coordinated but there are no differences after punishing.

While our primary focus is to compare the performance of coordinated and uncoordinated punishment, a relevant finding in the dyadic version of the investment game (Berg et al., 1995) is that allowing for punishment can have a detrimental effect on behavior. In our team investment game, we observe negative effects of allowing for punishment on joint investment or on the allocator's rate of return, but these effects appear only when punishment is uncoordinated. In this case, we observe the same behavior of investors and allocators when coordinated punishment involves increasing returns to scale compared to the treatment in which punishment is not possible. There are, however, differences in terms of efficiency: investors use punishment when it is available, which destroys the efficiency gains generated by coordinated punishment. In fact, when examining the total final payoffs for investors and allocators, the highest total payoffs are obtained when punishment is not possible.

To our knowledge, we are the first to investigate the effect of coordinated punishment in an asymmetric situation, such as the team investment game. There is an existing body of literature that systematically looks at the effects of

---

[2] In our setting with two investors, the need to coordinate the punishment decision implies that unanimity is required to inflict damage on the allocator.

punishment on the investment game (see, among others, Fehr & Rockenbach, 2003; Fehr & List, 2004; Houser et al., 2008; Rigdon, 2009; Calabuig et al., 2016). Yet these papers rely on the dyadic version of the game, thus they are not well-suited for comparing the effectiveness of uncoordinated and coordinated punishment in a team investment situation. The most closely related studies are the theoretical models of Boyd et al. (2010) and Olcina and Calabuig (2015), who highlight the benefits of coordinated punishment in an evolutionary setting. Boyd et al. (2010) consider a prisoners' dilemma and show that cooperation can be sustained as an equilibrium outcome when punishers divide the cost of the punishment, if they coordinate their actions and decide to punish. In the model of Olcina and Calabuig (2015), there are two investors and one allocator who interact in an overlapping-generations dynamic model. As in Boyd et al. (2010), it is possible to sustain a cooperative equilibrium in the presence of coordinated punishment when the (individual) cost of punishment decreases as the number of punishers increases. In their setting, however, there is also the possibility of peer punishment, since investors can punish each other after observing the punishment decision of the other team members. Our contribution to this literature is to directly compare the effects of coordinated and uncoordinated punishment in an asymmetric situation, such as the team investment game, which also resembles a hold-up team situation. In this vein, we show that coordinated punishment may be beneficial for the joint investment even if subjects do not divide the cost of the punishment and peer-punishment is not allowed.

The findings in our experiment dovetail with other studies that rely on the idea that punishment may require coordination, e.g., in the form of voting. Tyran and Felds (2006), Casari and Luini (2009), Ertan et al. (2009), Putterman et al. (2011), Noussair and Tan (2011) or Van Miltenburg et al. (2014), among others, allow subjects to vote over different punishment schemes in a public good game. Their results suggest that this usually results in efficiency gains because subjects tend to punish below-average contributors and strong cooperators are barely punished. Our findings, that investors tend to free-ride on the punishment decisions of others when punishment is uncoordinated but are more likely to punish together when punishment has to be coordinated, relate our paper to other studies that discuss the importance of conditional punishment in the public goods game (Casari & Luini, 2009; Cinyabuguma et al., 2006; Egas & Riedl, 2008; Kamei, 2014). Key to our discussion is the fundamental difference between the public goods game (where players have symmetric roles and identical opportunities to punish) and the team investment game (in which players differ in their roles and opportunities to punish). Furthermore, we do not allow for voting but, instead, consider that coordinated punishment occurs when players undertake an individual decision that is costly to them (e.g., workers decide to go on strike).

The rest of the paper is organized as follows. Section 2 presents our experimental design. We present our behavioral hypotheses in Sect. 3 and summarize our findings in Sect. 4. Section 5 concludes.

2 The team investment game with punishment

2.1 Experimental design

2.1.1 Investment game

We consider a team investment game with two investors and one allocator. In the game, each player is initially endowed with 20 Experimental Currency Units (ECUs, hereafter). They interact as follows:

- *STAGE 1 (Investment)*: Investors choose simultaneously whether or not to invest in a joint project. The investment decision is a binary decision that can be interpreted as investors putting effort or not into a joint project, or hiring an external agent. The individual cost of the investment equals 5 ECUs. When both investors decide to invest, the game proceeds to stage 2. If there is no joint investment, the game ends and the investor who decided to invest (if any) has to pay the cost of the investment.[3]
- *STAGE 2 (Return)*: The joint investment results in a surplus of 30 ECUs. In Stage 2, the allocator chooses the amount of ECUs to be returned to the investors. As the investment of each investor is equally valuable, we impose that any return $x \in [0, 30]$ must be equally divided between the two investors, so that each of them receives $x/2$. The allocator adds the amount he keeps, $(30 - x)$, to his initial endowment of 20 ECUs.
- *STAGE 3 (Punishment)*: Investors are allowed to punish the allocator upon observing the returned amount when punishment is possible. The punishment decision has an individual cost of 5 ECUs that is paid regardless of the punishment inflicted on the allocator. This decision can be interpreted as investors deciding whether or not to protest or go on strike. The allocator's payoffs are reduced in a given share $\lambda_n \in [0, 1]$, depending on the number of investors that choose to punish, $n \in \{0, 1, 2\}$. If none of the investors decide to punish, then no damage is inflicted on the allocator ($\lambda_0 = 0$). The reduction in the allocator's payoffs if only one of the investors punishes (λ_1) or if they both punish (λ_2) varies across treatments.

2.1.2 Treatments

In our $UP_{30,60}$ treatment, punishment is uncoordinated. If only one of the investors decides to punish, the allocator's payoffs are reduced by $\lambda_1 = 0.30$. If both investors decide to punish, the allocator's payoffs are reduced by $\lambda_2 = 0.60$.

---

[3] An example of investments that are complements in the workplace include the case in which workers' skills are complements. For instance, if one worker is an expert in IT (Information Technology) and the other worker is an expert in accounting and finance the project will be successful only if both workers provide effort (i.e., invest) in the project.

Table 1 Summary of treatment conditions (reduction in allocators' payoff (λ_n))

Treatment	None of the investors punish (λ_0)	Only one investor punishes (λ_1)	Both investors punish (λ_2)
$UP_{30,60}$	0%	30%	60%
$CP_{0,60}$	0%	0%	60%
$CP_{0,80}$	0%	0%	80%
NP	n.a	n.a	n.a

Note: Punishment has an individual cost of 5 ECUs and it is only allowed in Stage 3 if both investors decided to invest in Stage 1. In all the three treatments with punishment, investors have to pay the individual cost of punishment if they want to reduce the allocator's payoffs, regardless of whether or not the payoffs of the allocator are reduced. In our treatment with no punishment (NP) the game ends in Stage 2

In our $CP_{0,60}$ treatment**,** punishment has to be coordinated and the allocator's payoffs are reduced by $\lambda_1 = 0$ [$\lambda_2 = 0.60$] if one [both] of the investors decides [decide] to punish. Hence, this treatment incorporates the idea that *investors need to coordinate their actions* to reduce the allocator's payoffs.

Our $CP_{0,80}$ treatment incorporates the two characteristic aspects of coordinated punishment, i.e., the fact that coordination is required and the increasing returns are to scale. In this treatment, the allocator's payoffs are reduced by $\lambda_1 = 0$ [$\lambda_2 = 0.80$] if one [both] of the investors decides [decide] to punish.

Finally, we conduct a NP treatment with no punishment, i.e., there is no Stage 3 in which investors can punish the allocator after observing the return.

We summarize our treatment conditions in Table 1.

Payoffs: Let the dummy 1_i take the value 1 if investor i decides to invest and 0 otherwise, while the dummy 1_J stands for the case of joint investment, i.e., it takes the value 1 when both investors decide to invest, otherwise $1_J = 0$. Similarly, let the dummy 1_p denote whether investor i decides to punish or not. This variable takes the value 0 when punishment is not feasible, e.g., in the NP treatment.

The final payoff of each investor, π_i is determined as follows:

$$\pi_i = \begin{cases} 20 - 5 1_i & \text{if } 1_J = 0 \\ 15 + \left(\frac{x}{2}\right) - 5(1_p) & \text{if } 1_J = 1 \end{cases} \tag{1}$$

The final payoff for the allocator, π_A, is determined as follows:

$$\pi_A = \begin{cases} 20 & \text{if } 1_J = 0 \\ (50 - x)(1 - \lambda_n) & \text{if } 1_J = 1 \end{cases} \tag{2}$$

where $\lambda_n = \{\lambda_0, \lambda_1, \lambda_2\}$ denotes the reduction in the allocator's payoff after the punishment decisions (see Table 1).[4]

2.2 Procedures

We recruited a total of 270 subjects to participate in seven experimental sessions. There were 75 subjects (2 sessions) in each of the punishment treatments and 45 subjects (1 session) in our NP treatment.[5] All sessions were conducted at the LINEEX (University of Valencia). Subjects were undergraduate students with no experience in similar experiments. The experiment was conducted using the z-Tree software (Fischbacher, 2007), and no subject participated in more than one session. Subjects were recruited using the electronic recruitment system of the laboratory.

In our experiment, subjects played the team investment game for 15 periods in a partners-matching protocol with each subject's role, investor (Player A) or allocator (Player B), being fixed during the whole session.[6] In each period, investors had to choose whether or not to invest in a joint project. We employed the strategy method for allocators by asking them the amount of money that they would like to return if the investment turned out to be successful in each period. We decided to employ this method so as to have more observations for the case in which we could not observe joint investment. In addition, there is evidence that the strategy method does not affect the behavior of allocators in the investment game compared with the direct method (see the meta-analysis in Johnson & Mislin, 2011). Importantly, the decision of allocators was binding and disclosed to investors in the case of joint investment. More precisely, a screenshot at the end of each period informed subjects about the decisions of each investor in the group, the amount returned by the allocator (if there was joint investment) and the punishment decisions of investors when punishment was feasible. The type of feedback that subjects receive was kept constant across treatments, given that differences in information can alter behavior (e.g., see Nikiforakis, (2010) for evidence in public good games). We also kept constant the information that allocators received regarding the use of the strategy method, i.e., in none of

---

[4] It is worth noting that the value of λ_n can be interpreted as the *capacity of punishment* of investors, as this denotes how their punishment decisions affect the allocators' payoffs. In our setting, this capacity of punishment λ_n determines the share that investors can destroy from allocators and it is independent of the amount that allocators return. However, there is a direct relationship between the capacity of punishment of investors and the *fine-to-fee* ratio or the effectiveness of punishment, which can be defined as the factor by which punishment reduces the allocator's payoff (see Calabuig et al., 2016; 2022 for further discussion).

[5] This translates into 25 trios or independent observations in each of the punishment treatments (15 trios in the NP treatment).

[6] In the experiment, the repeated game was preceded by a practice round for subjects to get familiar with the software. Subjects were re-matched after the practice round (receiving no feedback) to play the repeated game, where we fixed the groups. At the end of the experiment, subjects were paid for their practice round and one randomly selected period of the repeated game. We observe no difference between the decisions in the practice round and the first period of the repeated game in any of the treatments. The interested reader can consult Appendix C for this analysis.

the treatments did the investors know that allocators were making a choice for the case in which there would be no joint investment.

All the amounts are referred to ECUs in our experiment, which were transformed into Euros to pay subjects (3 ECUS = 1€). On average, each subject received about 16€ for a 60-min session, including a 5€ show-up fee. At the end of the session, we collected information regarding the subjects' gender, age, cognitive reflection (Frederick, 2005), risk aversion (Gneezy & Potters, 1997) and trusting behavior (Glaeser et al., 2000). We use these variables as controls in our econometric analysis. Appendix B contains the translated version of the experimental instructions, the screenshots of the experiment and the complete questionnaire. This includes a summary of the demographic variables that we collected in our questionnaire, which do not differ across treatments (see Table B1).

2.3 Some (relevant) aspects of our team investment game

Before proceeding with our theoretical predictions, there are a few aspects of our experimental design that are worth discussing.

First, we extend the bilateral investment game in Berg et al. (1995) to the minimal setting in which coordinated punishment can be studied. Our setting has the advantage that we can derive tractable predictions in a theoretical model (see Appendix A). By studying coordinated punishment in a simplified environment, we can also obtain valuable insights into its underlying mechanisms and identify potential limitations. This can be useful when looking at the efficacy of coordinated punishment in more complex scenarios.

Second, our team investment game considers two investors and one allocator. This relates to the experimental paper by Cassar and Rigdon (2011), who vary the number of investors (1 or 2) and allocators (1 or 2) across treatments. There are at least two aspects that make our paper different from theirs. On the one hand, we allow for the possibility of punishment. On the other hand, we impose some symmetries in the game, e.g., we focus on the case in which investment decisions are complements as is the case in weak-link games (e.g., Brandts & Cooper, 2006; Harrison & Hirshleifer, 1989; Riedl et al., 2015; Van Huyck et al., 1990). This, in turn, implies that the allocator values the investment of each investor equally, thus the amount that the allocator returns to each investor will be the same (by design).[7]

Finally, since our examples on labor conditions can be associated with stable relationships, we decide to consider an experiment where subjects interact repeatedly using a fixed-matching protocol.

---

[7] Our goal is to compare both punishment devices when we avoid any interference from motivations such as envy or inequality aversion among investors, or any strategic behavior from allocators. In this vein, our design choices prevent investors making comparisons before deciding whether or not to punish the allocator. In fact, both investors are in the same position in the punishment stage, since joint investment is required to produce the surplus and the allocator cannot compensate only one of them (e.g., to reduce the likelihood of being punished if punishment has to be coordinated).

3 Theoretical predictions

Different behavioral models make different predictions, depending on the punishment device, if there is one. The self-interest model, for example, assumes that subjects are exclusively motivated by their material payoff. In this setting, costly punishment will never occur, regardless of the punishment device. Thus, selfish allocators will return nothing and, anticipating this, investors will not invest. This predicted outcome is not only inefficient but it is also in sharp contrast with the observed behavior in laboratory experiments (see Cooper and Kagel (2017) and Eckel and Wilson (2010) for a review of the literature and Johnson and Mislin (2011) for a meta-analysis).

There are other behavioral theories that provide more reasonable predictions on the effects of punishment on joint investments and return. In Sect. 3.1, we discuss the predictions when looking at coordinated and uncoordinated punishment. In Sect. 3.2, we discuss the predictions for the case in which we compare settings with and without punishment.

3.1 Coordinated versus uncoordinated punishment

Models of inequality aversion (e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) predict that punishment will affect the investor and the allocator's behavior in the team investment game, depending on the capacity and the cost of punishment. A key aspect in these theories is what constitutes a "fair" return. Inequality-averse investors would punish if they received less than the fair return and, anticipating this, the allocator will increase the return to avoid being punished.

In our team investment game, the punishment decision is a public good only when punishment is uncoordinated: any investor has to pay the cost of punishment, and the potential benefits (in terms of higher returns from allocators and lower inequality) will be enjoyed by both investors. Coordinated and uncoordinated punishment therefore differ in the "rules of the game" when a punishing subgame is reached. For selfish individuals this makes no difference in terms of behavior, but the strategic situation changes dramatically for individuals with social preferences who may be willing to punish unfair offers. As occurs in public good games (Fehr and Gächter2000, Carpenter, 2007, Egas & Riedl, 2008, Kamei, 2014), there is a social dilemma when punishment is uncoordinated, given that investors can free-ride on the punishment decision of others. This will result in lower levels of punishment, smaller returns from allocators and lower levels of joint investment. On the contrary, under coordinated punishment, the punishment subgame is a coordination game where the incentives to free-ride disappear, i.e., the requirement of coordination prevents the free-riding behavior of investors in the punishment stage.[8] In equilibrium, allocators anticipate that (for certain values of the cost of punishment)

---

[8] In fact, it is a step-level public good where a contribution threshold needs to be reached: a minimum number of players (threshold) are necessary to effectively punish the allocator.

investors will be more willing to punish if punishment has to be coordinated, thus they return a larger amount when punishment has to be coordinated than when it is uncoordinated.

In Appendix A, we formalize this idea by building on a two-sided incomplete-information model with two possible types of investors (selfish and inequality-averse) and two possible types of allocators (fair-minded and profit-maximizer). We prove the existence of some efficient pooling equilibria in which both selfish and inequality-averse investors decide to invest in equilibrium. We also show that if the share of inequality-averse investors and fair-minded allocators is sufficiently high, then, joint investment is more likely to occur when punishment requires coordination, and this effect is amplified for increasing returns to scale. A similar argument applies to allocators. If investors are more likely to punish when punishment has to be coordinated, allocators will have incentives to increase their return so as to avoid being punished. Hence, we expect higher returns from allocators when investors need to coordinate their actions.

Prediction 1 (Social preferences) *We expect to observe higher levels of joint investment and a higher return from allocators when punishment has to be coordinated, compared with the case in which punishment is uncoordinated. The effect amplifies when coordinated punishment displays increasing returns* ($UP_{30,60} < CP_{0,60} < CP_{0,80}$).

One important aspect in our model is that we expect for (inequality-averse) investors to free-ride on the punishment decision of others if punishment is uncoordinated. This cannot occur if punishment has to be coordinated because one investor cannot affect the level of inequality when punishing alone. Thus, we expect to observe higher levels of joint punishment when investors need to coordinate their actions.

Prediction 2 (Free-riding hypothesis) *We expect to observe higher levels of joint punishment when punishment has to be coordinated than when it is uncoordinated* ($UP_{30,60} < CP_{0,60} = CP_{0,80}$).

Regarding efficiency, there is no clear-cut prediction since Predictions 1 and 2 go in opposite directions. It will depend on which of the two outcomes (joint investment or joint punishment) plays a predominant role in efficiency.

3.2 Punishment versus non-punishment

How does punishment affect behavior compared to a situation in which there is no feasible punishment? The deterrence hypothesis in Gneezy and Rustichini (2000) posits that individuals are less willing to undertake an action if they will be punished. This, in turn, implies that allocators will return more if there is a threat of punishment, compared with a setting in which punishment is not possible. Anticipating this behavior, investors are more willing to invest if they can punish the allocator. Indeed, this is the prediction of our theoretical model in Appendix A.

Prediction 3 (Deterrence hypothesis) We expect to observe higher levels of joint investment and higher return from allocators when punishment is possible, compared with the case in which punishment is not possible.

While the deterrence hypothesis predicts a positive effect of punishment, evidence exists which shows that punishment may be detrimental for behavior, e.g., it might have a crowding-out effect on the intrinsic motivation of individuals. According to the models of intrinsic/extrinsic motivation, punishment have a direct "incentive effect" by altering subjects' extrinsic motivation, making investors more willing to invest and allocators more willing to reward. However, it may also have indirect and opposite psychological effects on the intrinsic motivation, crowding out the incentivized behavior; e.g., if punishment is possible and investors decide to invest, the allocator may intepret this behavior as a "bad signal" because investment decisions are not necessarilty linked to trust. This, in turn, can lower the intrinsic motivation of the investors to exert effort.[9]

Prediction 4 (Crowding-out effects) The levels of joint investment and the return from allocators will not necessarily increase when punishment is possible, compared with the case in which punishment is not possible.

The deterrence hypothesis predicts that the possibility of punishment will increase the levels of joint investment and the return by allocators, while models of intrinsic/extrinsic motivation that account for crowding-out effects allow for the opposite behavior. As a result, we would expect to observe $NP < UP_{30,60} < CP_{0,60} < CP_{0,80}$ if Predictions 1 and 3 are satisfied. If Predictions 1 and 4 are satisfied, then it would be possible to observe that NP lies in between any of the treatments where punishment is possible; it could even be the case that $UP_{30,60} < CP_{0,60} < CP_{0,80} \leq NP$, if allowing for punishment crowds out the intrinsic motivation to invest and return. In fact, the following prediction follows from Predictions 1 and 4:

Prediction 5 *Efficiency will not necessarily be higher when punishment is possible, compared with the case in which punishment is not possible.*

In our experiment, subjects play the team investment game for 15 periods in a partners-matching protocol with the role of each subject (investor or allocator) being fixed during the whole session. We rely on well-known results from finitely-repeated games with incomplete information to argue that our predictions do not change for this repeated version.[10]

---

[9] In fact, there is mixed evidence regarding the effect of punishment in the level of investment and the allocator's return in the dyadic version of the investment game (Calabuig et al., 2016; Fehr and List 2004; Fehr and Rockenbach 2003; Houser et al., 2008; Rigdon 2009). The work of Gneezy et al. (2011) or Xiao (2018) provide a comprenhesive review on the positive and negative effects that punishment can have on behavior.

[10] Using Folk theorems (Benoit and Krisna 1985, Friedman, 1971, Fudenberg and Maskin, 1986) we can posit that for a given finite horizon *T,* there is a set of distributions of the populations of investors and allocators for which there will be joint investment along most of the equilibrium path (except probably
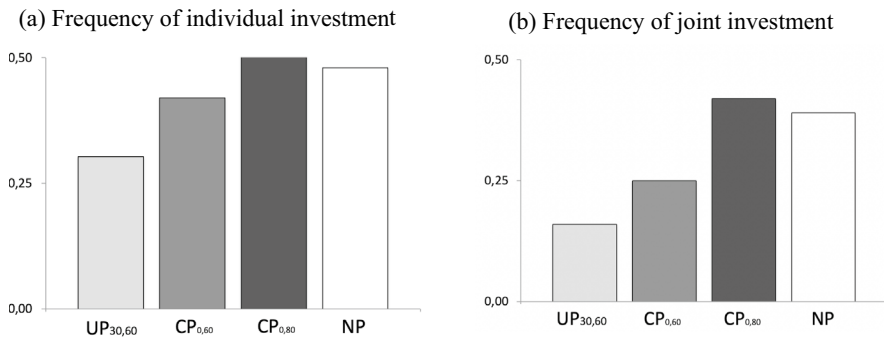
(a) Frequency of individual investment

(b) Frequency of joint investment

Fig. 1 Relative frequency of individual and joint investment in each treatment. Note: $UP_{30,60}$ refers to the uncoordinated punishment treatment, $CP_{0,60}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 60% if both investors punish and $CP_{0,80}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 80% if both investors punish. NP is the treatment in which there is no possibility to punish. In each treatment with punishment we have 50 investors and 25 allocators. We have 30 investors and 15 allocators in NP

4 Results

First, we analyze the levels of joint investment and the return set by allocators in Sect. 4.1. Then, we discuss the punishment behavior of investors and examine differences in efficiency across treatments in Sect. 4.2.

4.1 Investment decisions and returned amount

Section 4.1.1 presents some descriptive statistics and the results of our non-parametric analysisfor the level of the joint investment and the allocator's return. Section 4.1.2 includes the analysis of the joint investment and the intended return across periods. The main findings reported in these sections are robust to an econometric approach, as is discussed in Sect. 4.1.3 (the details of the econometric approach are relegated to Appendix E).

4.1.1 Descriptive statistics

Investment decisions. Figure 1a displays the relative frequency of individual investment in each treatment condition. The frequency of joint investment is shown in Fig. 1b.

Footnote 10 (continued)

for the last periods). In addition, it is also well-known that in a repeated game with finite horizon and if there is multiplicity of Perfect Bayesian Equilibrium in the stage game, any succession of Nash equilibria of the stage game will constitute a subgame perfect equilibrium of the repeated game (see Brown and Fehr. (2004) for a similar approach).

We focus our statistical analysis on the levels of joint investment in Fig. 1b, but the figures would be very similar if we analyzed the individual investment.[11] We find that the likelihood of joint investment is lower in $UP_{30,60}$ than in $CP_{0,60}$ (16% vs 25%, $p = 0.031$). When we look at the treatments with coordinated punishment, we observe that joint investment is lower in $CP_{0,60}$ than in $CP_{0,80}$ (25% vs 42%, $p = 0.017$). When we look at the behavior with and without punishment, we find a negative effect of allowing for punishment in that joint investment is more likely in NP than in $UP_{30,60}$ (39% vs 16%, $p = 0.003$). Similarly, there are differences between NP and $CP_{0,60}$ (39% vs 25%, $p = 0.063$). However, there are no differences in the likelihood of joint investment when comparing NP and $CP_{0,80}$ (39% vs 42%, $p = 0.97$).[12]

Result 1. (Predictions 1, 3 and 4) *Compared with uncoordinated punishment, joint investment is more likely if investors need to coordinate their punishment decisions. In addition, the increasing returns to scale amplifies the benefits of coordinated punishment. Finally, punishment reduces the likelihood of joint investment, except when investors need to coordinate their actions and there are increasing returns* $(UP_{30,60} < CP_{0,60} < CP_{0,80} = NP)$.

Intended return. As we use the strategy method, Fig. 2 displays the intended return by allocators in each treatment.[13][14]

We observe that allocators intend to return more in $CP_{0,60}$ than in $UP_{30,60}$ (12.1 vs 8.3, $p = 0.014$) but there are no differences between $CP_{0,60}$ and $CP_{0,80}$ (12.1 vs 11.6, $p = 0.72$). We also find that the intended return by allocators in the absence of punishment (NP: 9.90 ECUs) lies in between their intended return in the treatment with uncoordinated punishment ($UP_{30,60}$: 8.3 ECUs) and the intended return in treatments where punishment is coordinated ($CP_{0,60}$: 12.1 ECUs, $CP_{0,80}$: 11.6 ECUs). Our non-paramteric analysis indicates that differences in the return by allocators are (weakly) significant when we compare NP and $UP_{30,60}$ ($p = 0.087$), but there are no

---

[11] Unless otherwise noted, we rely on the Wilcoxon rank-sum (Mann–Whitney) test for pairwise comparisons. The *p*-values refer to one-tailed tests if we expect a directional alternative (e.g., higher levels of joint investment if punishment has to be coordinated). If there is no ex-ante expectation, the p-values refer to two-tailed tests. In our analysis, we pool the observations by groups across the 15 periods to guarantee independence. We do not compare $UP_{30,60}$ and $CP_{0,80}$ as two elements change across treatments, namely *i)* the need to coordinate the actions and *ii)* the increasing returns to scale (or the effectiveness of the punishment). The interested reader can consult Calabuig et al. (2016) for this analysis.

[12] If we correct the p-values for multiple testing hypothesis using Bonferroni-Holm the difference between NP and $CP_{0,60}$ becomes insignificant (corrected $p = 0.126$).

[13] Allocators do not know whether or not their choices will be implemented in a period; thus, one may argue that their decision takes place in a "cold" state. To address this issue, we may consider the return by allocators for periods of joint investment (effective return) or their return in a "hot" state, which may be assumed to occur when allocators observed joint investment in the previous period. We observe a similar pattern when considering these alternative measures (see Table D2, Appendix D).

[14] Figure D1 in Appendix D presents the distribution of intended return across treatments. We observe that allocators are heterogeneous in their return with spikes in the data taking place in 0 ECUs (no return), 10 ECUs (investors retrieve their investment), 15 ECUS (allocators return half of the surplus), and 20 ECUs (allocators divide the joint surplus equally among the three members of the team).
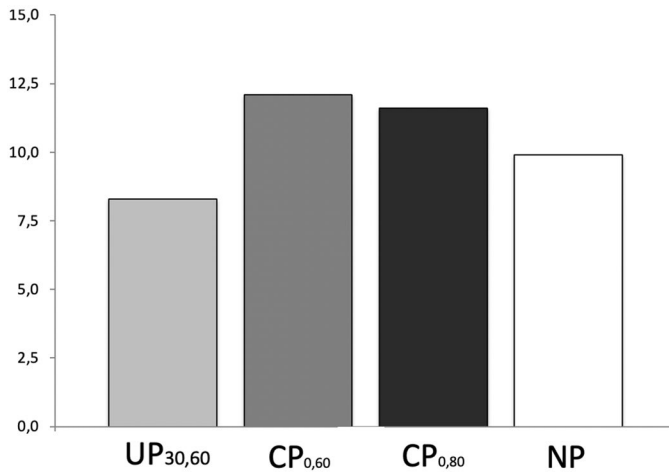
Fig. 2 Intended return by allocators in each treatment. Note: $UP_{30,60}$ refers to the uncoordinated punishment treatment, $CP_{0,60}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 60% if both investors punish and $CP_{0,80}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 80% if both investors punish. NP is the treatment in which there is no possibility to punish. In each treatment with punishment we have 50 investors and 25 allocators. We have 30 investors and 15 allocators in NP

differences when we compare the intended return in NP and any of the coordinated treatments ($p > 0.57$).[15]

Result 2. (Predictions 1, 3 and 4) *Compared with uncoordinated punishment, allocators return more if investors need to coordinate their actions, and the increasing returns to scale have no effect on the decision of allocators. In addition, the possibility of punishment reduces the return set by allocators, except when investors need to coordinate their actions* ($UP_{30,60} < CP_{0,60} = CP_{0,80} = NP$).

The two previous results support Prediction 1 and 4 (crowding-out effects) from our theoretical analysis. Prediction 3 is not supported. Our results suggest that coordinated punishment fosters the level of joint investment and boosts the reward set by allocators, compared with uncoordinated punishment. There is also evidence that allowing for punishment may be detrimental for the behavior of investors and allocators, but differences in the capacity of punishment are also crucial in explaining choices, in line with previous findings in the dyadic version of the investment game (Calabuig et al., 2016; Fehr & List, 2004; Fehr & Rockenbach, 2003; Houser et al., 2008; Rigdon, 2009).

---

[15] If we correct the p-values for multiple testing hypothesis using Bonferroni-Holm the difference between NP and $UP_{30,60}$ becomes insignificant (corrected $p = 0.34$).

Table 2 Investors' punishment decisions: Econometric analysis

	$UP_{30,60}$ vs $CP_{0,60}$		$CP_{0,60}$ vs $CP_{0,80}$	
	Logit	Heckman	Logit	Heckman
	(1)	(2)	(3)	(4)
$ReceivedReturn_t$	− 0.490***	− 0.069***	− 0.448***	− 0.060***
	(0.083)	(0.011)	(0.055)	(0.005)
$Punish_{t-1}$	− 0.290	0.0831	− 0.093	0.148***
	(0.268)	(0.073)	(0.228)	(0.0494)
$UP_{30,60}$ * other $punish_{t-1}$	− 0.356*	− 0.086***		
	(0.182)	(0.028)		
$CP_{0,60}$ * other $punish_{t-1}$	0.057	0.022	0.022	0.021
	(0.112)	(0.017)	(0.111)	(0.014)
$CP_{0,80}$ * other $punish_{t-1}$			0.260***	0.050***
			(0.097)	(0.012)
Period	− 0.053	0.001	− 0.060	− 0.005
	(0.049)	(0.010)	(0.040)	(0.005)
Coord. punish ($CP_{0,60}$)	1.555**	0.265***		
	(0.606)	(0.095)		
Coord. punish ($CP_{0,80}$)			− 0.201	− 0.058
			(0.556)	(0.060)
Constant	5.186**	1.166***	4.674	0.833***
	(2.482)	(0.347)	(2.916)	(0.280)
Heterogeneity	Yes	Yes	Yes	Yes
Number of obs	282	1,286	484	1,274

Note: The dependent variable is the decision to punish the allocator. *ReceivedReturn_t* is the amount that each investor receives from the allocator, *Punish_{t−1}* refers to the punishing decision of the investor in the preceding period, while *OtherPunish_{t−1}* refers to whether or not the other member of the team punished in the preceding period. We include the period in which decisions are made as well as dummy variables for the treatments

4.1.2 Dynamics

Dynamics of investment decisions: Figure 3a displays the frequency of individual investment across periods for each possible treatment. We observe a tendency for individual investment to decrease in every treatment, which is confirmed by the Jonckheere–Terpstra test ($p < 0.001$).[16] Relevant to our setting is the behavior of investors regarding the levels of joint investment in Fig. 3b. In this case, the Jonckheere–Terpstra test identifies decreasing trends in joint investment over periods

---

[16] We observe that the likelihood of individual investment varies significantly across treatments (Krusall Wallis, $p = 0.003$). When doing pairwise comparisons, the Wilcoxon rank-sum (Mann–Whitney) test suggests that investment is significantly higher when punishment has to be coordinated, as it is suggested by Result 1 above ($UP_{30,60}$ vs $CP_{0,60}$: $p = 0.07$; $CP_{0,60}$ vs $CP_{0,80}$: $p = 0.08$).

(a) Frequency of individual investment
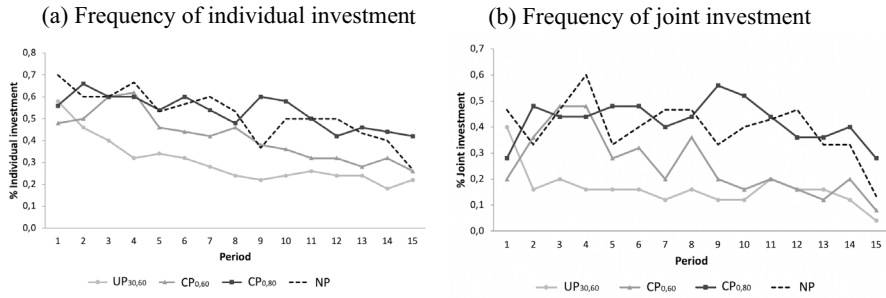
(b) Frequency of joint investment

Fig. 3 Relative frequency of joint investment across periods. Note: $UP_{30,60}$ refers to the uncoordinated punishment treatment, $CP_{0,60}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 60% if both investors punish and $CP_{0,80}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 80% if both investors punish. NP is the treatment in which there is no possibility to punish. In each treatment with punishment we have 50 investors and 25 allocators. We have 30 investors and 15 allocators in NP

in all the treatments NP ($p < 0.048$), $UP_{30,60}$($p = 0.009$), $CP_{0,60}$ ($p < 0.001$), except for the $CP_{0,80}$ ($p = 0.257$). Note that there is an end-time effect, especially in the NP treatment, where joint investment goes from 33% in period 14 to 13% in period 15. The results regarding the long-term trend hold if we remove period 15 from the analysis, except for the NP treatment (NP: $p = 0.438$, $UP_{30,60}$: $p = 0.039$, $CP_{0,60}$: $p = 0.001$, $CP_{0,80}$: $p = 0.89$).

Differences in the behavior of investors can also be identified at the group level. The frequency of groups that invest in a period, provided that they invested in the previous one, is 43% in $UP_{30,60}$, 62% in $CP_{0,60}$, 80% in $CP_{0,80}$, and 69% in NP. The results of a test of proportion suggest that groups who invest in one period in $CP_{0,60}$ are more likely to invest in the subsequent one compared with groups who invest in one period in $UP_{30,60}$ ($p = 0.028$). If we compare $CP_{0,60}$ and $CP_{0,80}$ the results suggest that the increasing returns to scale amplifies this effect ($p = 0.029$). As for the comparison with the NP treatment, the frequency of groups that keep investing after doing so in one period is 69%. This figure is statistically different from 43% in the $UP_{30,60}$ treatment ($p = 0.009$). The differences are not significant when we compare NP with any of the treatments with coordinated punishment $CP_{0,60}$ (69% vs 62%, $p = 0.48$) and $CP_{0,80}$ (69% vs 80%, $p = 0.23$) (see Appendix D1 for other results regarding group heterogeneity).

Dynamics of the intended return: Figure 4 displays the intended return by allocators across periods for each possible treatment. There are at least three results regarding the intended return that are worth mentioning. First, we cannot detect any trend using the Jonckheere–Terpstra test (NP: $p = 0.522$, $UP_{30,60}$: $p = 0.287$, $CP_{0,60}$: $p = 0.169$, $CP_{0,80}$: $p = 0.397$), thereby suggesting that the behavior of allocators is quite stable across periods. Second, there is an end-time effect in all treatments, which seems to be pronounced in the NP treatment where investors cannot punish the allocator. In this setting, the intended return goes from 9.67 ECUs in period 14 to 2.33 ECUs in period 15.
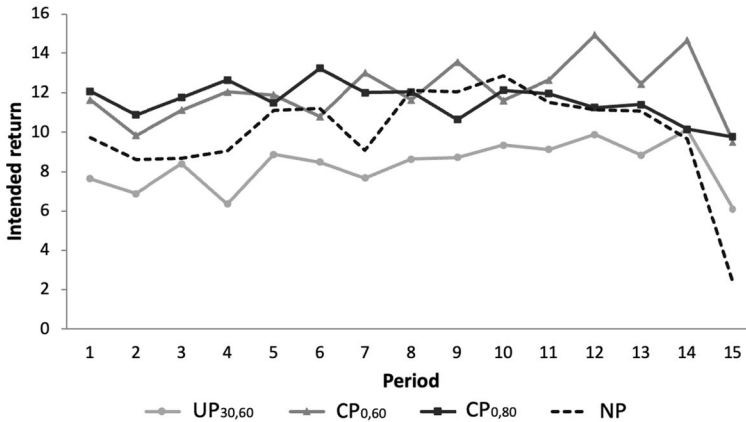
Fig. 4 Intended return by allocators across periods. Note: $UP_{30,60}$ refers to the uncoordinated punishment treatment, $CP_{0,60}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 60% if both investors punish and $CP_{0,80}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 80% if both investors punish. NP is the treatment in which there is no possibility to punish. In each treatment with punishment we have 50 investors and 25 allocators. We have 30 investors and 15 allocators in NP

Finally, we observe that the intended return by allocators is below the horizontal line of 10 ECUs when punishment is uncoordinated. This is important because some authors relate reciprocity to the idea that the investment decision should be repaid so that it is worth engaging in some form of cooperation in the future (Gambetta, 2000, Alos-Ferrer & Farolfi, 2019). In the $UP_{30,60}$ treatment, investors do not retrieve on average what they invested, which may discourage further investment (bear in mind that there is a decrease over time in the joint investment in $UP_{30,60}$).

4.1.3 Econometric analysis

Our previous findings are in line with our Prediction 1, that coordinated punishment is beneficial to joint investment and return levels, compared with uncoordinated punishment. Moreover, we find that the increasing returns to scale might help in encouraging higher levels of joint investment, while it barely has any effect on the allocators, i.e., it is enough for investors to coordinate their actions to get an increase in the return set by allocators. There are some variables (e.g., the individual characteristics or the history of decisions) that we have not taken into account in our previous analysis and could affect our results. In order to isolate their effect, we present an econometric analysis in Appendix E, where we study the determinants of the joint investment and the intended return using Arellano–Bond (1991) and random effect (hereafter, RE) specifications.[17] To control for individual heterogeneity, we

---

[17] We believe that Arellano-Bond (1991) is appropriate to our setting since we have a potential endogeneity problem (due to the partners matching) and we do not have exogenous variables to use as instruments; in fact, a test of serial autocorrelation confirms that the endogeneity problem is present in our data. See Roodman (2006, 2009) for a discussion on how to select a valid set of instruments in the Arel-

consider the variables: subjects' gender, age, cognitive reflection, risk aversion and trusting behavior (see Appendix B for the corresponding definitions). In our analysis, we take into account the different dynamics across treatments by controlling for the period in which investors and allocators make their choices. In what follows, we summarize the main findings of our econometric approach.

Joint investment: We confirm the positive effects of coordinated punishment on the levels of joint investment, compared with uncoordinated punishment, e.g., the dummy variable $CP_{0,60}$ is always significant when comparing it with $UP_{30,60}$ (Arellano–Bond: $p < 0.005$, RE logit: $p < 0.049$). The difference between $CP_{0,60}$ and $CP_{0,80}$ is also statistically significant at any common significance level (Arellano–Bond: $p < 0.001$, RE logit: $p < 0.044$). When we compare the investors' behavior when punishment is possible and when it is not, we confirm that investors are less likely to invest in $UP_{30,60}$ compared withNP (Arellano–Bond: $p < 0.056$, RE logit: $p < 0.075$). The same holds when comparing the $CP_{0,60}$ treatment with NP (Arellano–Bond: $p < 0.066$, RE logit: $p < 0.081$). There are no significant differences in the behavior of investors in the $CP_{0,80}$ and NP treatments (Arellano–Bond: $p > 0.254$, RE logit: $p > 0.490$). In addition, our estimates for the level of joint investment indicate that there is evidence of "homegrown trusting preferences" as defined by Cassar and Rigdon (2011), i.e., investors are more likely to invest if they did so in the previous period.[18]

Intended return: Our econometric approach confirms that coordinated punishment boosts the returned amount, compared with uncoordinated punishment. This is explained by the fact that investors need to coordinate their actions, rather than by the increasing returns to scale, i.e., the intended return is larger in $CP_{0,60}$ than in $UP_{30,60}$ (Arellano–Bond: $p < 0.001$, RE Tobit: $p < 0.048$) but it is indistinguishable in $CP_{0,60}$ and $CP_{0,80}$ (Arellano–Bond: $p > 0.103$, RE Tobit: $p > 0.77$). As for the effect of allowing for punishment, we find that allocators return less in $UP_{30,60}$ compared to NP (Arellano–Bond: $p < 0.005$, RE Tobit: $p < 0.032$). When punishment has to be coordinated, the Arellano-Bond specification suggests that allocators return more compared to the case in which punishment is not possible, while the random-effect Tobit specification suggests that there are no significant differences in the behavior of allocators in the NP treatment and the treatments with coordinated punishment $CP_{0,60}$ and $CP_{0,80}$(Arellano–Bond: $p < 0.008$, RE Tobit: $p > 0.484$ for $CP_{0,60}$ vs NP; Arellano–Bond: $p < 0.003$, RE Tobit: $p > 0.412$ for $CP_{0,80}$ vs NP). As a result, we can conclude that allowing for punishment is detrimental to the behavior of allocators, except when punishment has to be coordinated.

Footnote 17 (continued)

lano-Bond model. For other experimental papers that use this methodology see Fischbacher and Gächter (2010), Brañas-Garza et al. (2013) or Charness et al. (2017), among others.

[18] One interesting finding from the Arellano-Bond model is that using punishment facilitates the joint investment when punishment has to be coordinated, i.e., observing that both investors punished in the previous period has a negative (positive) effect on the likelihood of joint investment when punishment is uncoordinated (coordinated) (see Table E1 in Appendix E). We discuss this issue further in Sect. 4.2 below where we focus on examining the differences in punishment behavior across treatments.
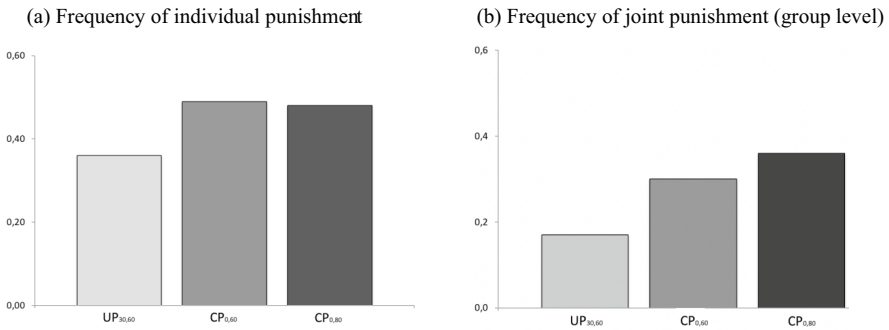
(a) Frequency of individual punishment

(b) Frequency of joint punishment (group level)

Fig. 5 Punishment behavior in each treatment. Note: $UP_{30,60}$ refers to the uncoordinated punishment treatment, $CP_{0,60}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 60% if both investors punish and $CP_{0,80}$ refers to the coordinated treatment in which the amount reduced from the allocators' payoffs is 80% if both investors punish. The number of observations in each treatment for the individual punishment is 50 and for the joint punishment is 25

4.2 Punishment behavior and efficiency

So far, we have shown that coordinated punishment can facilitate investor's decisions on investment and foster the allocators' return, compared with uncoordinated punishment. However, punishment decisions can also harm efficiency (Calabuig et al., 2016; Chaudhuri, 2011). In Sect. 4.2.1 we look into the punishing behavior of investors in each treatment. We then examine in Sect. 4.2.2 the effect of these decisions on the sum of total payoffs, which we take as a measure of efficiency.

4.2.1 Punishment behavior

Figure 5 summarizes the punishment behavior of investors in each treatment when punishment is feasible. Figure 5a shows that investors are more likely to punish when punishment has to be coordinated. The differences are (weakly) significant using a test of proportion if we compare $UP_{30,60}$ with$CP_{0,60}$ (36% vs 49%, $p = 0.088$). The differences are not statistically significant for treatments with coordinated punishment (49% vs 48%, $p = 0.92$). Note that joint punishment is required to inflict some damage on the allocator if punishment has to be coordinated. In Fig. 5b we observe that joint punishment is less likely in $UP_{30,60}$ than in $CP_{0,60}$ (17% vs 31%, $p = 0.051$), while differences are not significant when we compare $CP_{0,60}$ and $CP_{0,80}$ (31% vs 36%, $p = 0.702$).[19]

---

[19] The interested reader on the dynamics of punishment can consult Appendix D (Figures D2 and D3). When we look at the likelihood of punishment we observe no clear trend in $UP_{30,60}$ or $CP_{0,60}$, using the Jonckheere-Terpstra test, but there is a decreasing trend in $CP_{0,80}$ ($p < 0.001$).

Result 3 (Prediction 2) Investors are more likely to punish if punishment has to be coordinated; in fact, joint punishment is more likely to be seen in $CP_{0,60}$ and $CP_{0,80}$ than in $UP_{30,60}$.

The previous result supports Prediction 2 that the levels of joint punishment are higher when punishment has to be coordinated. Next, we attempt to explain the determinants of punishment decisions by means of an econometric analysis in Table 2. It should be remembered that punishment cannot be observed in our team investment game unless there is joint investment. Thus, our first specification (Logit) estimates the individual decision to punish only for cases in which there was joint investment; otherwise, we have missing values of the dependent variable. Our second specification relies on the two-step method in Heckman (1979) to estimate the likelihood that the investor will punish the allocator.[20] In our model, the willingness to punish depends on the return received from the allocator ($ReceivedReturn_t$) and the punishment decisions of investors. Our first variable ($Punish_{t-1}$) refers to the punishing decision of the investor in the preceding period, which may be important if subjects have expressive preferences for punishment, as noted in Casari and Luini (2012). Kamei (2014) shows that punishment decisions can also be conditional on others' punishment decisions. Our explanatory variable $OtherPunish_{t-1}$ considers the punishment decision of the other member of the team in the preceding period. This variable is important in our setting as we expect investors (not) to *free-ride* if punishment is uncoordinated (coordinated), respectively. Thus, we interact this variable with the treatment conditions. All regressions control for individual heterogeneity; in particular we consider the subjects' gender, age, cognitive reflection, risk aversion and trusting behavior as measured by Glaeser et al. (2000) (see Appendix B for the corresponding definitions).

As expected, the likelihood to punish decreases with the allocators' return. We also observe that investors are less likely to punish if their partner did inflict punishment in the preceding period, when punishment is uncoordinated. Such *free-riding* behavior is not observed in the coordinated devices; in fact, our analysis for $CP_{0,80}$ suggests that investors are more likely to punish if their partner did so in the preceding period.

Result 4 (Prediction 2) In the $UP_{30,60}$ treatment, investors are less likely to punish if they observe that their partner did punish in the preceding period (free-riding behavior). In the $CP_{0,80}$ treatment, investors are more likely to punish if they observe that their partner did punish in the preceding period.

---

[20] The sample-selection model in Heckman (1979) estimates first a probit model on the probability of the dependent variable to be observed. In our setting, this refers to the probability of observing joint investment (see Table E4 in Appendix E for this model). Then, a regression of maximum likelihood with the subsample is considered, including the Heckman's lambda (obtained in the first step) as an additional regression. Our findings are robust to other specifications (see Calabuig et al., 2022).
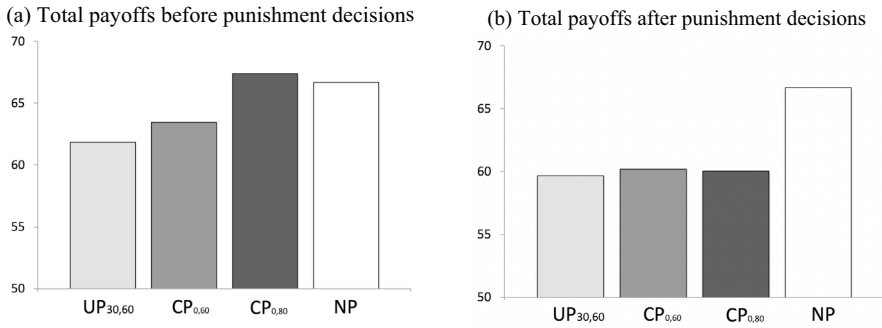
(a) Total payoffs before punishment decisions   (b) Total payoffs after punishment decisions

Fig. 6 Efficiency across treatments

Overall, our findings suggest that the higher level of joint investments and the higher returns from allocators when punishment has to be coordinated (Results 1 and 2) are accompanied by higher levels of joint punishment and less free-riding behavior (Results 3 and 4).

4.2.2 Efficiency

In this section, we first look at the average reduction in allocators' payoffs and then, we examine how differences in punishing behavior translate into differences in efficiency, which we measure using the sum of total payoffs. Not surprisingly, the average loss in allocators' payoffs due to punishment is the highest in $CP_{0,80}$ (12.45 ECUs). In this treatment, both the frequency of joint punishment and the reduction in the allocators' payoffs when both investors decide to punish (λ_2) are the highest. While we observe less frequency of joint punishment in $UP_{30,60}$ than $CP_{0,60}$, the reduction in allocators' payoffs is higher in the former treatment ($UP_{30,60}$: 9.82 ECUs, $CP_{0,60}$: 7.88 ECUs). This (partially) occurs because allocators return more when punishment has to be coordinated, so that the reduction is lower in $CP_{0,60}$.[21] Figure 6 shows the sum of total payoffs both (a) before and (b) after the investors' punishment decision.

The main finding that can be gleaned from Fig. 6 is that all benefits of coordinated punishment disappear once we account for the investors' punishment decision. In fact, we find that the sum of total payoffs after punishment is roughly the same across treatments where punishment is possible ($p > 0.366$ for pairwise comparisons). If we look at the effect of punishment on behavior, we find that NP outperforms all other treatments in terms of efficiency when we look at the final payoffs (NP: 66.67 ECUs, $UP_{30,60}$: 59.67 ECUs, $CP_{0,60}$: 60.20, $CP_{0,80}$: 60.05 ECUs). Statistically, differences are significant when we compare the sum of total payoffs in

---

[21] We find no significant differences when comparing the reduction in allocators' payoffs in $UP_{30,60}$ and $CP_{0,60}$ ($p = 0.235$) or when comparing $CP_{0,60}$ and $CP_{0,80}$ ($p = 0.270$).

NP with the one in the rest of treatments with punishment (($p < 0.005$ for pairwise comparisons). We summarize these findings as follows:

Result 5 (Prediction 5) Efficiency is the highest when punishment is not feasible. If there is punishment, there are no differences in efficiency between coordinated and uncoordinated punishment, once we account for the punishment decision by investors.

This result provides evidence in favor of Prediction 5 that punishment does not lead to efficiency gains, compared with the case in which punishment is not possible. In the Appendix D we report the investors' and the allocators' payoffs before and after the punishment decisions (see Table D4). Our data show that punishment decisions mainly affect the allocators' payoffs.

5 Concluding remarks

Coordinated punishment is a prevalent phenomenon in society as sometimes punishers need to coordinate their actions and "punish together" for the punishment to be successful. In this paper, we look at the effects of coordinated punishment in an asymmetric situation that resembles the labor relationship. While there is mounting evidence on the effects of punishment in various settings, we are not aware of any paper that directly examines the comparison between coordinated and uncoordinated punishment in an asymmetric situation, such as the team investment game. Our paper is an attempt to fill this gap.

In our experimental setting we assume that investors may need to coordinate their actions and punish together to inflict any damage to the allocator. In addition, we allow for coordinated punishment to exhibit returns to scale, compared with uncoordinated punishment. We find that joint investment is more likely when punishment has to be coordinated. We also find that allocators return more to investors when punishment has to be coordinated. However, these findings do not translate into higher levels of efficiency: we observe that total payoffs are the same under uncoordinated and coordinated punishment, probably because the higher levels of joint investment and returns observed in the latter device are associated with a higher tendency for investors to *punish together* when punishment has to be coordinated. Thus, the more frequent use of joint punishment under the coordinated condition (Prediction 2) can be compensated by the benefits (higher investment and return) of coordinated punishment (Prediction 1).

In our paper, we also consider a setting in which punishment is not possible. In this respect, our findings seem to indicate that the negative effects of punishment that have been identified in the dyadic version of the investment game (Calabuig et al., 2016; Fehr & List, 2004; Fehr & Rockenbach, 2003; Houser et al., 2008) are also present in our team investment game, especially when punishment is uncoordinated. These negative effects on the levels of joint investment and the allocators' return tend to vanish when punishment has to be coordinated, but are

present when we look at the levels of efficiency: the sum of total payoffs is always higher if there is no possibility to punish (Prediction 5). Overall, these findings suggest that models that incorporate crowding-out effects (Prediction 4) might fit the data better than those that rely on the deterrence hypothesis (Prediction 3).

Our paper might be viewed as a first attempt to compare the effects of coordinated and uncoordinated punishment in an asymmetric situation when players have different roles and opportunities to punish. One important question to be addressed concerns how (and why) coordinated punishment emerges in a society or in a group. This is a relevant question as some authors have shown that it may be difficult to explain the evolution of cooperation from an evolutionary perspective when punishment is uncoordinated (Bowles et al., 2012; Boyd et al., 2010; Guala, 2012). We argue that punishment decisions may be considered as a public good. In this regard, coordinated punishment, as opposed to uncoordinated punishment, has the great advantage of eliminating the free-riding behavior of punishers. This idea paves the way to rationalizing the benefits of coordinated punishment in a team hold-up relationship.

We believe that there are other aspects of coordinated punishment that may be worth considering in future research. An interesting extension would be to consider a continuum of investors. Our design could also incorporate the possibility of communication (Charness & Dufwenberg, 2006; Choi & Lee, 2014), given that punishment is sometimes "coordinated by means of gossip and other communications" (Boyd et al., 2010; Fehr & Williams, 2013). While these features would require using other analytical tools, the results of these models could shed light on the success of mobilizations and protests that may be required for punishment to be coordinated, e.g., the social upheavals in Chile or the Arab Spring, among others. Noticeably, extending the results to these settings may also result in considering a different game, e.g., we may need to use threshold models in which people observe the action of others (Bikhchandani et al., 1992, 1998; Lohmann, 2000). In this regard, revolutions have been modeled as coordination or stag-hunt games in De Mesquita (2010), Edmond (2013), Kiss et al. (2017) or Barbera and Jackson (2020), and some models assume that protesters may have information on what others have done (e.g., Barbera & Jackson, 2020; Kiss et al., 2017). In this setting, communication constraints can affect the possibility of coordinated punishment, e.g., it may be more costly to coordinate actions when punishment has to be coordinated; thus one may wonder when and under what conditions this punishment device can thrive.

There are other aspects of our design that can be extended as well. For example, we deliberately focus on the case in which investment is a binary decision and the returns of the allocator are equally split between the investors. We have in mind a labor setting in which investors (or workers) choose whether or not to exert effort in a common project that requires complementarities and their effort is equally valuable for the allocator (i.e., the firm). A natural extension would be to consider a setting in which investors can choose different investments and the allocator is allowed to reward them differently (Cassar & Rigdon, 2011). It may also be possible to consider a setting in which the creation of the surplus does not require the joint investment, but it is suffice that one investor decides to invest to generate efficiency gains.

Last, but not least, it seems reasonable to wonder whether people would opt for an institution with coordinated punishment if they had the opportunity to choose,in line with our paper, where subjects can endogenously decide the punishment institution they want to implement, if any (Fehr & Williams, 2013; Kosfeld et al., 2009). We observe that workers become members of trade unions so as to negotiate agreements on pay and conditions with their employers, but are they really aware of the positive effects of being united?

We believe that testing this question is another avenue for future research, as it may be relevant to explain the (natural) emergence of institutions where punishment requires coordination by the group members. Overall, these possible extensions reflect important situations that have not yet been studied. We hope our research sparks further interest in these areas.

Declarations

Conflict of interest The authors declare that they have no affiliations with or involvement in any organization or entity with any financialor non-financial interest in the subject matter or materials discussed in this paper.

References

Alós-Ferrer, C., & Farolfi, F. (2019). Trust games and beyond. *Frontiers in Neuroscience, 13*, 887.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*, 277–297.

Barbera, S., & Jackson, M. (2020). A model of protests, revolution, and information. *Quarterly Journal of Political Science, 15*(3), 297–335.

Baron, J. N., & Kreps, D. M. (1999). *Strategic Human Resources: Frameworks for General Managers*. Wiley.

Bejarano, H., Gillet, J., & Rodriguez-Lara, I. (2020). Trust and Trustworthiness After Negative Random Shocks. ESI Working Paper 20–25.

Benoit, J., & Krishna, V. (1985). Finitely repeated games. *Econometrica, 53*(4), 905–922.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Gamesand Economic Behavior, 10*, 122–142.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*(5), 992–1026.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives, 12*(3), 151–170.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*, 166–193.

Bowles, S., Boyd, R., Mathew, S., & Richerson, P. J. (2012). The punishment that sustains cooperation is often coordinated and costly. *Behavioral and Brain Sciences, 35*, 20–21.

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science, 328*(5978), 617–620.

Brañas-Garza, P., Buchelia, M., Espinosa, M. P., & García-Muñoz, T. (2013). Moral cleansing and moral licenses: Experimental evidence. *Economics and Philosophy, 29*(2), 199–212.

Brandts, J., & Cooper, D. J. (2006). A change would do you good. An experimental study on how to overcome coordination failure in organizations. *American Economic Review, 96*(3), 669–693.

Brown, F., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica, 72*(3), 747–780.

Calabuig, V., Fatas, E., Olcina, G., & Rodriguez-Lara, I. (2016). Carry a big stick, or no stick at all: Punishment and endowment heterogeneity in the trust game. *Journal of Economic Psychology, 57*, 153–171.

Calabuig, V., Jimenez, N., Olcina, G., & Rodriguez-Lara, I. (2022). United We Stand: On the Benefits of Coordinated Punishment. ESI Working Paper 22–12.

Carpenter, J. P. (2007). The demand for punishment. *Journal of Economic Behavior and Organization, 62*(4), 522–542.

Casari, M., & Luini, L. (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior and Organization, 71*(2), 273–282.

Casari, M., & Luini, L. (2012). Peer punishment in teams: Expressive or instrumental choice? *Experimental Economics, 15*(2), 241–259.

Cassar, A., & Rigdon, M. (2011). Trust and trustworthiness in networked exchange. *Games and Economic Behavior, 71*(2), 282–303.

Charness, G., Cobo-Reyes, R., Jiménez, N., Lacomba, J. A., & Lagos, F. (2017). Job security and long-term investment: An experimental analysis. *European Economic Review, 95*, 195–214.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74*(6), 1579–1601.

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

Choi, S., & Lee, J. (2014). Communication, coordination, and networks. *Journal of the European Economic Association, 12*(1), 223–247.

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

Ciriolo, E. (2007). Inequity aversion and trustees' reciprocity in the trust game. *European Journal of Political Economy, 23*, 1007–1024.

Cooper, D., & Kagel, J. (2017). Other-regarding preferences: A selective survey of experimental results. In J. H. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics* (pp. 217–289). Princeton.

De Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review, 104*(3), 446–466.

Eckel, C., & Wilson, R. (2010). Trust and social exchange. In J. Druckman, D. Green, J. Kuklinski, & A. Lupia (Eds.), *Handbook of Experimental Political Science* (pp. 243–257). Cambridge University Press.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies, 80*(4), 1422–1458.

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society b: Biological Sciences, 275*(1637), 871–878.

Ertan, A., Page, T., & Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review, 53*(5), 495–511.

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

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives-trust and trustworthiness among CEOs. *Journal of the European Economic Association, 2*, 743–771.

Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature, 422*(6928), 137–140.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics, 114*, 817–868.

Fehr, E., & Williams, T. (2013). *Endogenous emergence of institutions to sustain cooperation*. Mimeo.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Fischbacher, U., & Gachter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review, 100*(1), 541–556.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Friedman, J. (1971). A non-cooperative equilibrium for supergames". *Review of Economic Studies, 38*(1), 1–12.

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica, 54*(3), 533–554.

Gambetta, D. (2000). *"Can we trust trust", making and breaking cooperative relations, electronic edition* (pp. 213–237). University of Oxford.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics, 115*(3), 811–846.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*(4), 191–209.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics, 112*(2), 631–645.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies, 29*(1), 1–17.

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences, 35*, 1–59.

Harrison, G. W., & Hirshleifer, J. Z. (1989). An experimental evaluation of weakest link best shot model of public goods. *Journal of Political Economy, 97*, 201–225.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47*, 153–161.

Houser, D., Xiao, E., McCabe, K., & Smith, V. (2008). When punishment fails: Research on sanctions, intentions and non-cooperation. *Games and Economic Behavior, 62*(2), 509–532.

Johnson, N., & Mislin, A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889.

Kamei, K. (2014). Conditional punishment. *Economics Letters, 124*(2), 199–202.

Kiss, H. J., Rodriguez-Lara, I., & Rosa-Garcia, A. (2017). Overthrowing the dictator: a game-theoretic approach to revolutions and media. *Social Choice and Welfare, 49*, 329–355.

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

Lohmann, S. (2000). Collective action cascades: An informational rationale for the power in numbers. *Journal of Economic Surveys, 14*(5), 655–684.

Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior, 68*(2), 689–702.

Noussair, C. N., & Tan, F. (2011). Voting on punishment systems within a heterogeneous group. *Journal of Public Economic Theory, 13*(5), 661–693.

Olcina, G., & Calabuig, V. (2015). Coordinated punishment and the evolution of cooperation. *Journal of Public Economic Theory, 17*, 147–173.

Putterman, L., Tyran, J.-R., & Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics, 95*, 1213–1222.

Riedl, A., Rohde, I. M., & Strobel, M. (2015). Efficient coordination in weakest-link games. *Review of Economic Studies, 83*(2), 737–767.

Rigdon, M. (2009). Trust and reciprocity in incentive contracting. *Journal of Economic Behavior and Organization, 70*, 93–105.

Roodman, D. (2006). How to do xtabond2: An Introduction to "Difference" and "System" GMM in Stata. Center for Global Development. Working paper 103.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*(1), 135–158.

Smith, A. (2011). Income inequality in the trust game. *Economics Letters, 111*, 54–56.

Tyran, J., & Feld, L. (2006). Achieving compliance when legal sanctions are nondeterrent. *Scandinavian Journal of Economics, 108*(1), 135–156.

Van Huyck, J., Battalio, R., & Beil, R. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review, 80*(1), 234–248.

Van Miltenburg, N., Buskens, V., Barrera, D., & Raub, W. (2014). Implementing punishment and reward in the public goods game: The effect of individual and collective decision rules. *International Journal of the Commons, 8*(1), 47–78.

Xiao, E. (2018). Punishment, social norms, and cooperation. *Research Handbook on Behavioral Law and Economics* (pp. 155–173). Edward Elgar Publishing.