palgrave communications
HUMANITIES | SOCIAL SCIENCES | BUSINESS

ARTICLE

https://doi.org/10.1057/s41599-019-0249-2 OPEN

Democracy matters: a psychological perspective on the beneficial impact of democratic punishment systems in social dilemmas

Rebekka Kesberg[1] & Stefan Pfattheicher[2]

ABSTRACT The implementation of punishment has proven a prominent solution to prevent the breakdown of cooperation in social dilemma situations. In fact, numerous studies show that punishment possibilities are effective in maintaining cooperative behavior. However, punishment is often not efficient in terms (a) of monetary benefits and in light of the fact (b) that punishment of cooperators (i.e., antisocial punishment) can occur. Still, recent research revealed that individuals vote for the implementation of such punishment systems. We address this contradiction by proposing that the benefits of democratic punishment systems in particular cannot be solely captured by monetary outcomes. Instead, the implementation of democratic punishment systems may enhance the *psychological benefits* of justice perceptions, satisfaction, and trust. Using iterated public goods games, the findings of the present study reveal not only higher cooperation levels and total payoffs in two different democratic punishment systems compared to other systems, but also higher justice perception, satisfaction, and trust. Furthermore, participants indicated the highest willingness to continue interactions in democratic punishment systems. Moreover, satisfaction, not monetary outcomes, was the best predictor of participants' willingness to stay in a system. Therefore, we argue that the efficiency of democratic punishment systems cannot be measured solely in monetary outcomes but that psychological benefits must be considered.

[1] Faculty of Engineering, Computer Science and Psychology Institute of Psychology and Education, Social Psychology, Albert-Einstein-Allee 47, 89069 Ulm, Germany. [2] Department of Psychology and Behavioural Sciences, Bartholins Allé 11, 8000 Aarhus, Denmark. Correspondence and requests for materials should be addressed to R.K. (email: rebekka.kesberg@uni-ulm.de)

Introduction

In a globalized world where many collective problems can only be solved in a cooperative manner (e.g., climate change actions), studies investigating the possibilities and limitations of human cooperation are undoubtedly more important than ever (Camerer, 2003; Rand and Nowak, 2013; Van Lange et al., 2013). Explaining cooperation in anonymous settings is particularly intriguing because cooperation is often costly for an individual, while the benefits of cooperation are created on the collective level rather than on the individual level. This divergence between individual and collective interests results in a social dilemma of conflicting interests (Hardin, 1968; Van Lange et al., 2013).

For decades researchers from various disciplines have investigated the (evolutionary) origins of cooperative behavior (e.g., Rand et al., 2012), contextual factors like group size (Barcelo and Capraro, 2015) and time pressure (Rand et al., 2012; Tinghög et al., 2013), and individual differences related to cooperative behavior including—but not limited to—motivation and social preferences (e.g., social value orientation, Balliet et al., 2009; Bogaert et al., 2008), personality (e.g., Hilbig et al., 2012; Capraro et al., 2014), trust (Van Lange et al., 1998), and morality (Capraro and Rand, 2018).

Interestingly, it seems that even though research on cooperation in social dilemmas is plentiful, most studies still evaluate successful cooperation and institutional designs to promote cooperation (like sanctions) through *monetary* or material outcomes (Balliet et al., 2011). There is, however, remarkably little research examining how other factors alone and in combination with monetary outcomes (on a group and individual level) relate to cooperative behavior, how they relate to the evaluation of institutional designs, and whether they predict future willingness to interact in the same institution. The current article aims to address this research gap and sheds light on psychological factors, i.e., satisfaction, justice perceptions, and trust, which emerge from and relate to cooperative and destructive behavior in social dilemmas due to differences in the institutional design of punishment systems.

Thus, compared to previous work in the field, our research does not focus on which individuals or which traits cooperative individuals inherit. Rather we examine in which systems most individuals are inclined to cooperate. Therefore, our research focuses on institutional factors, not individual factors, that foster cooperative behavior and reduce destructive behavior in social dilemmas. We investigate how different punishment systems, which are one possibility for establishing cooperation (e.g., Balliet et al., 2011), are beneficial in a psychological sense by enhancing justice perceptions, trust, and satisfaction. Moreover, we examine if these factors are advantageous for the establishment and maintenance of cooperation, irrespective of monetary outcomes. We test our hypothesis by implementing a standard public goods game (Fehr and Gächter, 2002) and four modified versions with peer punishment options. By examining these psychological constructs, we aim to obtain a more comprehensive picture of institutional factors that facilitate the emergence and maintenance of cooperation and then to examine which system is most beneficial for society. First we offer a brief overview of social dilemma research focusing on punishment systems, then describe psychological factors related to decision making.

Cooperation is often necessary in social dilemma situations in which individuals achieve the highest profit by being uncooperative, however, if everybody is uncooperative the whole group is worse off (Dawes, 1980). Research findings suggest that in these situations the establishment of cooperation is difficult because it is costly to the individual, and thus people are tempted to free-ride (Kim and Walker, 1984). One of the most prominent proposed solutions to overcoming social dilemmas is the establishment of a peer punishment system, i.e., a system in which each individual can sanction each of their group members. Studies have consistently shown that peer punishment systems stabilize (high) cooperation rates over time and that individuals become more cooperative in future interactions after being sanctioned (Fehr and Gächter, 2002; Guala, 2012).

These beneficial effects of peer punishment are, however, overshadowed by several harmful and unwanted side-effects of peer punishment. First, peer punishment is costly to the individual thus leading to a second-order social dilemma (Ostrom, 1990). Individually everybody is better off if they do not invest money to punish other group members; however, the whole group is worse off if uncooperative individuals are not sanctioned for their behavior. Second, the monetary costs of punishment lead to a lower net benefit on both individual and group levels, i.e., the profits from a high contribution rate do not cover the punishment costs (Dreber et al., 2008; Egas and Riedl, 2008; Fehr and Gächter, 2002; Gächter et al., 2008). Third, sanctions in a peer punishment are (often) not explicitly restricted to sanctioning uncooperative individuals; thus, antisocial punishment can emerge (i.e., the sanctioning of cooperative individuals) (Herrmann et al., 2008; Rand and Nowak, 2011). This behavior entails two unfavorable consequences: (a) again, costly punishment further reduces the net benefit on individual and group levels, and (b) cooperative but punished individuals are less likely to cooperate in future interactions (Herrmann et al., 2008).

To overcome the disadvantages of peer punishment systems, researchers examined centralized punishment systems. Centralized punishment (i.e., only higher authorities or institutions are allowed to punish uncooperative individuals) makes it impossible for counter punishment to occur, thus preventing a breakdown of cooperation due to fear of retaliation (Nikiforakis and Normann, 2008). Furthermore, when a higher authority was implemented, groups achieved higher payoffs and higher cooperation rates (Hilbe et al., 2014).

In addition to centralized punishment systems, researchers also investigated the effects of voting decisions on cooperation (2006). A study by Hauser et al. (2014) showed that voting systems enable cooperative individuals to restrain egoistic tendencies of other group members and thus prevent overexploitation of scarce resources. In another study by Ertan et al. (2009) a majority of individuals voted for the implementation of a punishment system that only allows punishment of low contributors. Ambrus et al. (2017) found higher cooperation rates and net payoffs if individuals could vote directly for or against the execution of a fixed punishment amount of another individual. Pfattheicher et al. (2018) replicated their findings by implementing a democratic peer punishment system in which individuals could not only vote for or against the execution of punishment of another individual, but the strength of the punishment was also determined by the group members. Their findings demonstrate one major advantage of voting systems compared to peer punishment systems: Cooperative individuals are less likely to be punished, that is, there is less executed antisocial punishment. Thus, the net benefit on the group level was higher in the democratic punishment system compared to the peer punishment system. The higher cooperation rates did not, however, cover the punishment costs, thus the highest net benefit emerged in the system without punishment options.

Taken together, these findings pose an interesting question: Why are people willing to implement a punishment system if the net benefits are lower compared to a system without punishment? What benefits do people hope to achieve by establishing punishment? One reason could be that people do not consciously realize that the monetary payoff is lower. However, one could also hypothesize that the perceived benefits of a punishment system are not

solely monetary; rather, any kind of punishment system might fulfill some other psychological needs which are important to establish and maintain cooperative behavior, for example, enhancing trust between group members. The current article aims to investigate these psychological factors in order to obtain a better understanding of other possible benefits of democratic punishment systems, i.e., satisfaction, perceived justice, and trust.

To make and evaluate decisions in ambiguous situations like social dilemmas, people often rely on their subjective experiences of, among others features, justice in a system (Van den Bos and Lind, 2002). Based on previous findings in the context of work and industrial psychology, there are four central concepts that relate to prosocial behavior. People often rely on these concepts when judging behavior and decisions in the social dilemma context: (1) procedural justice, (2) distributive justice, (3) satisfaction, and (4) trust (Cohnen-Charash and Spector, 2001). To follow is a conceptual overview of these concepts.

Justice is a commonly used term and covers a broad range of concepts including fairness. Although in previous literature justice and fairness have often been used to describe the same underlying concept (e.g., Rabin, 1993; Fehr and Schmidt, 1999), it is possible to distinguish between justice and fairness on a conceptual level. While fairness refers to equal opportunities to be advantaged, justice in general refers to just rewards of merit (Schurter and Wilson, 2009). In many social dilemma situations, everyone has the opportunity to take advantage of the others (e.g., free-riding), so based on that definition social dilemmas can often be labeled as fair. It is more difficult to determine whether or not they are just. Moreover, to fully understand the effects of justice perceptions on cooperation and punishment, it is important to distinguish two different kinds of justice: *distributive justice* and *procedural justice*.

Distributive justice refers to the perceived justice of outcome allocations, while procedural justice refers not to the outcome allocation itself, but instead to the process by which the outcome allocation was determined.

In the field of organizational and work psychology in particular, the relations of justice perceptions with outer constructs have been extensively studied. Justice perceptions were among others related to trust in and commitment to the organization (Colquitt et al., 2006; Moorman, 1991), job satisfaction (Pillai et al., 2001), and to organizational citizenship behavior (OCB; Bies et al., 1993; Brockner et al., 1987). Overall, findings consistently show that both procedural and distributive justice are associated with a broad range of behaviors which promote cooperation, e.g., OCB.

In line with the Equity Theory (Messick and Sentis, 1983), justice perceptions in an interdependent situation can be based on the social comparison of individual outcomes in relation to individual inputs (i.e., distributive justice perceptions) and on the process through which outcomes are determined (i.e., procedural justice). Comparing the process, procedural justice perceptions should be high when people have the opportunity to influence the process and when the process is transparent. Considering the comparisons of input–outcome relation, distributive justice perceptions should be high irrespective of the payoff as long as everybody receives what s/he deserves in relation to the input.

It seems plausible that even if distributive justice is given, other important psychological constructs might differ as a function of the payoff, i.e., satisfaction and trust. In fact, justice perceptions can be seen as predecessors of satisfaction since people rely on them to refer their satisfaction to an outcome (Van den Bos, Lind et al., 1997; Van den Bos, Vermunt et al. 1997). Mostly, people rely on distributive justice perceptions to refer their outcome satisfaction. However, Van den Bos and colleagues (1997) showed that without information about the outcome of others (i.e., no information about the distributive justice), participants judge their own outcomes based on their perception of the process (i.e., procedural justice), and thus referred their outcome satisfaction. A study by Van den Bos (1999) revealed that distributive justice and satisfaction do not automatically go hand in hand. While equal outcomes are typically perceived as more just than unequal outcomes, satisfaction is influenced by the amount of the outcome (e.g., a higher outcome is better than a lower one) and, in the case of unequal outcomes, the direction of the inequality (i.e., advantageous versus disadvantageous inequality). In general, people perceive advantageous inequalities (i.e., own outcome is higher than the outcome of others) as more satisfying than disadvantageous inequalities (i.e., own outcome is lower than the outcome of others). Bringing these concepts into the context of social dilemmas, we assume that people rely on distributive, as well as procedural justice to judge their satisfaction with their individual outcomes.

In an interdependent setting trust can either be defined as a belief about others' benevolent motives during a social interaction (e.g., Balliet and Van Lange, 2013a; Barber, 1983; Simpson, 2007; Yamagishi, 2011) or as the ability to predict others' behavior (e.g., Dasgupta, 1988; McAllister, 1995; Sitkin and Roth, 1993; Zucker, 1986). Studies investigating the relationship between dispositional trust and cooperative behavior are plentiful, and in general they show a positive relation between trust and cooperation (Balliet and Van Lange, 2013b; Deutsch, 1960). In addition, trust is also related to punishment efficacy, i.e., punishment is especially useful to enhance cooperative behavior in high trust societies Balliet and Van Lange, 2013b). Interestingly, a study by Mulder et al. (2006) revealed a more complex relationship between cooperation, punishment, and trust. Their findings revealed that sanctions might enhance external trust (i.e., trust that others cooperate to avoid sanctions) and reduce internal trust (i.e., trust that other cooperate due to an intrinsic motivation). However, only when participants first experienced a sanction system and then a system without sanctions did a negative effect on cooperation rates occur. In sum, these findings indicate that sanctions have a positive effect on external trust, that is, enhancing the predictability of others' behavior and thus enhancing cooperation.

Trust has also been related to justice perceptions, especially procedural justice. Without information about the trustworthiness of others, people rely on perceived procedural justice to judge their outcome (Van den Bos and Lind, 2002). Taken together, the reported findings support the assumption that cooperative behavior is related to justice perceptions which in turn promote trust and satisfaction.

Overall, the purpose of the present study was to obtain a better understanding about how characteristics of different punishment systems relate to individuals' perception of justice, satisfaction, and trust and how they influence cooperative behavior in a social dilemma. Moreover, the aim of the current study was to investigate not only the relationship between system characteristics, individuals' perception, cooperative behavior, and monetary outcomes, but also how these characteristics relate to participants' willingness to stay in a given system.

Our methodical procedure is based on the study by Pfattheicher et al. (2018). That is, we implemented a system without punishment opportunities, a system with peer punishment, and a system with democratic punishment (see Table 1). Our first hypotheses focus on replicating their findings and strengthening the claim about the beneficial aspects of democratic punishment systems over an extended interaction period:

> H1: Contribution rates to the public good in systems with punishment will be higher than contribution rates in the system without punishment.

Table 1 Overview over the different punishment systems

System	Punishment possible?	Voting possible?	Shared cost	Individual payoff per period
System I: Public goods game without punishment	×	×	×	Endowment—contribution + ¼ of the common good
System II: Public goods game with peer punishment (individual costs)	✓	×	×	Endowment—contribution + ¼ of the common good—own MUs invested to sanction others—received punishment
System III: Public goods game with peer punishment (shared costs)	✓	×	×	Endowment—contribution + ¼ of the common good—1/3 MUs invested to sanction others—received punishment
System IV: Public goods game with democratic punishment (individual costs)	✓	✓	✓	Endowment—contribution + ¼ of the common good—own MUs invested to sanction others if executed—received executed punishment
System V: Public goods game with democratic punishment (shared costs)	✓	✓	✓	Endowment—contribution + ¼ of the common good—1/3 MUs invested to sanction others if executed—received executed punishment

H2: Antisocial punishment will be higher in the peer punishment system compared to the democratic punishment system.

H3: The group net benefit in the peer punishment system will be lower compared to the democratic punishment system and the system without punishment.

H4: Satisfaction, trust, and perceived justice (procedural and distributive) will be higher in the democratic punishment system compared to the peer punishment system and the system without punishment.

Furthermore, we extend previous findings by implementing two modified punishment systems, i.e., a modified peer punishment system and a modified democratic punishment system. Contrary to the systems by Pfattheicher et al. (2018), in these new systems the costs of punishment are split equally among group members. Like in most societies, the cost of sanctioning an uncooperative individual are covered by the whole society (i.e., the cost for imprisonment is covered by tax money). Therefore, everybody's outcome is influenced by punishment decisions.

We assume that in a peer punishment system with shared punishment costs uncontrollable punishment will evolve; antisocial punishment cannot be prevented and additionally, individuals who punish antisocially can even engage in antisocial punishment without high costs for themselves. As a result, individuals have less control over their own outcome, and we therefore assume that the perceived procedural and distributive justice will be lower compared to all other systems. Likewise, trust and satisfaction, as well as individual and group net benefits should be lower compared to all other systems.

H5: Antisocial punishment in the peer punishment system with shared costs will be higher compared to all other systems.

H6: The group net benefit in the peer punishment system with shared costs will be lower compared to all other systems.

H7: Satisfaction, trust, and perceived justice (procedural and distributive) will be lower in the peer punishment system with shared costs compared to all other systems.

In a democratic punishment system with shared punishment costs, we assume that less antisocial punishment will emerge compared to all other systems. As every group member is affected by a punishment decision, we hope to overcome the second-order dilemma, and thus only 'necessary' punishment will be executed. Additionally, the perceived procedural and distributive justice should be higher compared to all other systems, because each group member can actively influence the decisions and is affected by them. Thus, antisocial punishment should be reduced, leading to higher individual and group net benefits compared to other punishment systems. As a result, satisfaction and trust should be higher than compared to all other systems.

H8: Antisocial punishment in the democratic punishment system with shared costs will be lower compared to all other systems.

H9: The group net benefit in the democratic punishment system with shared costs will be higher compared to all other systems.

H10: Satisfaction, trust, and perceived justice (procedural and distributive) will be higher in the democratic punishment system with shared costs compared to all other systems.

Finally, we assume that people will prefer the democratic punishment system with shared punishment cost over all other systems. In that system, individuals have the highest participation possibility which should enhance experienced justice, satisfaction, and trust, irrespective of the sole monetary outcome, and thus lead to a preference of the democratic systems. This means that we assume that psychological factors will be more influential on the willingness to stay in a system than the pure monetary outcomes.

H11: Willingness to continue in the democratic punishment system with shared cost will be higher compared to all other systems.

H12: Psychological constructs are stronger predictors for participants' willingness to continue than total payoff.

Method

Sample and procedure. In total, 76 Germans ($M_{age} = 22.92$; 49 women) participated in the study. To ensure anonymous conditions participants were seated in separate cubicles. All interactions were computer-mediated using z-Tree (Fischbacher, 2007). After signing an informed consent, participants first read the

instruction of all five game versions. Thus, before playing the games, participants were fully informed about the duration of the group interaction and that there would be five different versions.

For reasons of statistical power, we chose a within-person design, which means all participants ran through all five versions in random order (randomization was done on group level).[1] A power analysis (G*Power; Faul et al., 2009) revealed very high statistical power ($P = 0.95$; alpha level = 0.05, two-tailed, correlation among repeated measurements = 0.30) to detect a main effect between the five game versions of small-to-medium size ($f = 0.15$) (Cohen, 1992). At the end of the session, participants were anonymously paid their earnings ($M = €11.78–\$13.44$, $SD = 1.75$).

Public goods game. We applied five different versions of a standard public goods game (Fehr and Gächter, 2002) to test our hypotheses concerning the effects of different sanction systems on justice perceptions, satisfaction, trust, and willingness to stay in a given system. All versions are displayed in Table 1.

Our basic experimental setup followed the design by Pfattheicher and colleagues (2018). Overall, participants played 50 periods of the public goods games, i.e., ten periods per version. Participants were told that they would remain in the same group of four players across the 50 periods. But to prevent possible carry-over effects between periods, after every period, the group was randomly shuffled. Thus, the participants remained in the same group of four players, but the players' numbers were randomly changed (e.g., a participant played as Player #1 in the first period then played as Player #3 in the second period).

Each group consisted of four players, and each player was endowed with 20 monetary units (MUs; 1 MU was equal to 1€ cent–1.12US$ cent). Players were free to choose how many MUs to keep and how many to contribute to the public good. The total sum of the public good was multiplied by 1.6 and then equally split between the group members. That is, each player received one-fourth of the public good, independent of his or her individual contribution. Accordingly, it was always in the material self-interest of every individual to keep all of the MUs, irrespective of how much the other three subjects contributed to the group project. Following the contribution, each player was given accurate information about the contributions made by the other three players, the payoff from the public good, their period, and total payoff. At this point, the game variant without punishment opportunities ends.

In the game variant with the peer punishment system, a punishment stage was added after the feedback stage with information about each player's contribution and their payoff from the public good. In that stage, participants could invest their own MUs (between zero and ten for each player) to punish their group members. The invested MUs reduced payoff of the other player by a factor of three (e.g., the investment of two MUs decreases the payoff of another by six MUs). After these decisions, each player was given information about how much they had punished, how much they have been punished, their period, and total payoff.

In the democratic peer punishment system, additional decisions were added to the peer punishment system. After the punishment stage, participants were presented with information about each player's contribution, their payoff from the public good, and how much all players in total intend to punish each other group member. Then players voted for or against the execution of the intended punishment for every other player (by clicking "yes" or "no"). Punishment was executed if a simple majority of group members (i.e., two out of three) voted for the punishment of a fourth player. If there was a majority, all players had to bear the costs of their own intended punishment. Finally, each player was given accurate information about how much they punished, whether there was a majority of group members that voted for the player to be punished, how much they have been punished, and their period and total payoff.

In the systems with shared costs, participants could also invest their own MUs, but each player had to cover one-third of the costs irrespective of his or her own investment. For example, if Player 1 invested ten MUs to reduce Player 4 payoff, while Player 2 invested zero MUs and Player 3 invested two MUs, then the overall costs of the punishment are 12. The costs were split equally between Players 1, 2, and 3, i.e., everyone had to pay four MUs. In the democratic system, participants only had to cover the cost when a majority voted for the execution of the punishment. An overview of the different systems is displayed in Table 1.

Punishment calculation. For each participant, we calculated two different punishment scores, i.e., how much he or she engaged in altruistic and in antisocial punishment. Punishment directed at group members who contributed *less* than the participant him/herself was classified as altruistic punishment (Fehr and Gächter, 2002). Conversely, punishment directed at group members who contributed *as much or more* than the participant him/herself was classified as antisocial punishment (Herrmann et al., 2008).

Psychological measures. After each game variant, participants indicated their perceived procedural and distributive justice, their satisfaction, their trust in other group members, and their willingness to continue in the given system. To assess procedural justice, three items of a scale by Folger and Konovsky (1989) were adapted to the present context. The items read, "In the last ten rounds I had the opportunity to influence the outcome of my group"; "In the last ten rounds personal motives (e.g., revenge) of other group members influenced my individual payoff"; and, "In the last ten rounds the process which determined the reduction of income was fair." The scale endpoints of all items were labeled (1) *not at all true* and (7) *completely true*. Initial reliability analysis revealed a very poor internal consistence. After omitting the second item, scale reliability in each game version was better and ranged from Cronbach's α's > 0.35 (democratic punishment with individual costs) to Cronbach's α's > 0.68 (peer punishment with shared costs), but still poor. Due to the game structure, in the game version without punishment opportunities procedural justice was only measured with the item "In the last ten rounds I had the opportunity to influence the outcome of my group." Therefore, based on theoretical considerations, we decided to include the first and the third item separately in the analyses. In fact, procedural justice is a very broad construct, and both items might capture different aspects. The first captures the overall possibility to participate, while the third item captures the perceived justice of the reduction process, irrespective of participation possibilities.

To assess distributive justice, three items of a scale by Prince and Mueller (1986) adapted to the context. The items read, "My income in the last ten rounds was just considering my contributions."; "My income in the last ten rounds was just compared to the contributions of my other group members."; and, "In the last ten rounds the decisions about the reduction of the income of other group members was just." The reliability of the perceived distributive justice items in each game version was very good (Cronbach's α's > 0.75). Due to the game structure, in the game version without punishment opportunities the item "In the last ten rounds the decisions about the reduction of the income of other group members was just." was excluded, thus the distributive justice measure only consisted of two items.

To assess trust, three items on basis of the Short Interpersonal Trust Scale (Beierlein et al., 2012) were adapted to the context. The items were chosen to capture how predictable the behavior of other group members was in each system. The items read, "In the last ten periods, the behavior of my group members was predictable"; "In the last ten periods, I was familiar with the behavior of my group members"; and, "In the last ten periods, the behavior of my group members was consistent." The reliability of the items assessing trust in each game version was very good (Cronbach's α's > 0.89).

To assess satisfaction, three items of a scale by Fitzsimons et al. (1997) were adapted to the context. The items read, "Overall, I am satisfied with my experiences in the last ten periods"; "Based on my experiences, I would repeat my decisions in the last ten rounds"; and, "In the last ten periods, I was satisfied with the system." The reliability of the items assessing system satisfaction in each game version was very good (Cronbach's α's > 0.77).

Results

We conducted mixed model ANOVAs comparing simultaneously all five systems to test our hypothesis. For clarity, we report our results in the same sequences as we displayed our hypotheses. An overview of all central findings is displayed in Tables 2 and 3. A more fine-tuned analysis focusing on cooperation levels and period payoff across the ten periods also replicated previous findings (Pfattheicher et al., 2018) can be found in the supplementary materials.

Replication of previous findings (H1–H4)

Contribution. In line with our assumption and previous findings (Fehr and Gächter, 2002; Pfattheicher et al., 2018), average cooperation rates in systems with punishment were significantly higher compared to the system without punishment (H1). Focusing on the difference in cooperation levels between the different punishment options, cooperation levels further increased in the democratic punishment system compared to the peer punishment system.

Punishment. Overall, there was no significant difference in the intended punishment between the peer punishment and the democratic punishment system. However, there was a significant difference in the executed punishment. In line with our hypothesis, this was driven by more executed antisocial punishment directed towards cooperative individuals in the peer punishment system (H2).

Total payoff. In line with previous findings, in the peer punishment system the higher cooperation levels did not cover the costs of punishment (Dreber et al., 2008; Egas and Riedl, 2008; Fehr and Gächter, 2002). However, the total payoff was significantly higher in the democratic punishment system compared to the peer punishment system and did not significantly differ from the system without punishment.

Psychological constructs. Highest levels of procedural and distributive justice, satisfaction, and interpersonal trust could be observed in the systems with the highest possible participation, that is, the democratic punishment system. Distributive justice was significantly lower in the peer punishment system and the system without punishment options compared to democratic punishment system. Perceived participation possibility (i.e., Procedural Justice 1) was significantly lower in the system without punishment options compared to all other systems. In addition, the peer punishment system did not differ from the democratic punishment system.

Table 2 Overview about the central findings across game conditions including mean and standard deviation

	Game condition					Overall statistic	
	No punishment	Peer punishment with individual costs	Peer punishment with shared costs	Democratic peer punishment with individual costs	Democratic peer punishment with shared costs	Pillai's trace	*F*-value
Contribution (in MUs)	85.26 (60.33)^{a}	131.00 (58.19)^{b}	137.34 (57.75)^{b,c}	144.20 (53.56)^{c}	140.22 (55.73)^{c}	0.49	17.33***
Intended punishment (in MUs)	n.a.	20.92 (36.26)^{a,b}	27.37 (50.75)^{a}	18.58 (37.14)^{b}	22.93 (41.51)^{a,b}	0.07	1.89
Punishment of uncooperative individuals	n.a.	8.70 (11.36)	9.16 (10.20)	10.17 (19.43)	10.64 (17.22)	0.02	0.37
Antisocial punishment	n.a.	11.86 (34.06)^{a,b}	18.21 (47.02)^{a}	8.41 (21.47)^{b}	12.29 (34.28)^{b}	0.07	1.81
Executed punishment (in MUs)	n.a.	20.92 (36.26)^{a}	27.37 (50.75)^{a}	9.68 (20.48)^{b}	11.36 (17.95)^{b}	0.22	6.98***
Punishment of uncooperative individuals	n.a.	8.70 (11.36)	9.16 (10.20)	6.56 (11.64)	7.96 (13.54)	0.05	1.14
Antisocial punishment	n.a.	11.86 (34.06)^{a}	18.21 (47.02)^{a}	3.12 (10.51)^{b}	3.39 (10.91)^{b}	0.13	3.63**
Total Income (in MUs)	251.16 (40.17)^{a}	194.92 (114.77)^{b}	197.44 (82.99)^{b}	246.83 (66.68)^{a}	238.13 (76.77)^{a}	0.50	17.80***
Procedural Justice 1	4.05 (2.25)^{a}	5.03 (1.90)^{b}	4.89 (1.97)^{b}	5.28 (1.64)^{b}	5.17 (1.81)^{b}	0.20	4.57**
Procedural Justice 2	n.a.	4.21 (1.81)^{a}	4.01 (2.04)^{a}	5.14 (1.70)^{b}	5.01 (1.72)^{b}	0.29	9.80***
Distributive Justice	4.23 (1.78)^{a,b}	4.69 (1.53)^{a}	4.29 (1.79)^{b}	5.12 (1.45)^{c}	5.20 (1.56)^{c}	0.27	6.76***
Trust	4.86 (1.68)^{a,b}	4.80 (1.59)^{a}	4.68 (1.72)^{a}	5.24 (1.63)^{b}	5.29 (1.57)^{b}	0.17	3.75***
Satisfaction	5.18 (1.56)^{a,c}	4.94 (1.45)^{a}	4.54 (1.73)^{b}	5.59 (1.22)^{c}	5.51 (1.26)^{c}	0.28	6.89***
Continue in system	4.37 (2.20)^{a,c}	4.03 (1.99)^{a,b}	3.62 (2.22)^{b}	5.04 (1.89)^{c}	4.84 (2.17)^{c}	0.28	7.07***

Note: All psychological variables were measured on a scale from *(1) not at all true* to (2) *completely true*. Groups with different letters (a–e) are significantly different on a 5% significance level
**p < 0.01; ***p < 0.00

Table 3 Individual comparisons

		Individual comparisons									
		I: NP vs. PPIC	II: NP vs. PPSC	III: NP vs. DPIC	IV: NP vs. DPSC	V: PPIC vs. PPSC	VI: PPIC vs. DPIC	VII: PPIC vs. DPSC	VIII: PPSC vs. DPIC	IX: PPSC vs. DPSC	X: DPIC vs. DPSC
Contribution (in MUs)	*t*(75)	−7.17	−7.21	−8.48	−7.33	−1.37	−3.31	−2.11	−1.60	−0.67	0.93
	p	0.000	0.000	0.000	0.000	0.173	0.001	0.038	0.113	0.506	0.354
	Cohens *d*	0.82	0.83	0.97	0.84	0.16	0.38	0.24	0.18	0.08	0.11
Intended Punishment (in MUs)	*t*(75)	n.a.	n.a.	n.a.	n.a.	−1.52	0.89	−0.55	2.30	1.44	−1.23
	p					0.133	0.378	0.584	0.024	0.154	0.224
	Cohens *d*					0.17	0.10	0.06	0.26	0.17	0.14
Punishmentt of uncooperative individuals	*t*(75)	n.a.	n.a.	n.a.	n.a.	−0.29	−0.65	−0.89	−0.52	−0.85	−0.19
	p					0.770	0.520	0.377	0.605	0.398	0.850
	Cohens *d*					0.03	0.07	0.10	0.06	0.10	0.02
Antisocial punishment	*t*(75)	n.a.	n.a.	n.a.	n.a.	−1.64	1.12	−0.14	2.32	2.03	−1.36
	p					0.106	0.268	0.892	0.023	0.046	0.179
	Cohens *d*					0.19	0.13	0.02	0.27	0.23	0.16
Executed punishment (in MUs)	*t*(75)	n.a.	n.a.	n.a.	n.a.	−1.52	3.94	2.69	3.86	3.25	−0.073
	p					0.133	0.000	0.009	0.000	0.002	0.466
	Cohens *d*					0.17	0.45	0.31	0.44	0.37	0.08
Punishment of uncooperative individuals	*t*(75)	n.a.	n.a.	n.a.	n.a.	−0.29	1.20	0.39	1.83	0.75	−0.78
	p					0.770	0.236	.692	0.071	0.455	0.441
	Cohens *d*					0.03	0.14	0.05	0.21	0.09	0.09
Antisocial punishment	*t*(75)	n.a.	n.a.	n.a.	n.a.	−1.64	2.65	2.55	3.16	3.18	−0.24
	p					0.106	0.010	.013	0.002	0.002	0.812
	Cohens *d*					0.19	0.30	0.29	0.36	0.37	0.03
Total Income (in MUs)	*t*(75)	4.70	5.40	0.61	1.49	−0.24	−6.51	−3.66	−6.29	−4.40	1.06
	p	0.000	0.000	0.546	0.141	0.811	0.000	0.000	0.000	0.000	0.293
	Cohens *d*	0.54	0.62	0.07	0.17	0.03	0.75	0.42	0.72	0.50	0.12
Procedural Justice 1	*t*(75)	−3.72	−2.85	−4.14	−3.74	0.57	−1.12	−0.72	−1.50	−1.43	0.47
	p	0.000	0.006	0.000	0.000	0.572	0.266	0.473	0.138	0.257	0.637
	Cohens *d*	0.43	0.33	0.48	0.43	0.07	0.13	0.08	0.17	0.13	0.05
Procedural Justice 2	*t*(75)	n.a.	n.a.	n.a.	n.a.	0.96	−4.27	−3.51	−5.03	−3.69	0.31
	p					0.341	0.000	0.001	0.000	0.000	0.543
	Cohens *d*					0.11	0.49	0.41	0.58	0.42	0.07
Distributive Justice	*t*(75)	−1.70	−0.19	−3.31	−3.39	2.11	−2.17	−2.44	−4.04	−4.18	−0.39
	p	0.093	0.845	0.001	0.001	0.038	0.033	0.017	0.000	0.000	0.698
	Cohens *d*	0.20	0.02	0.38	0.39	0.24	0.25	0.28	0.46	0.48	0.04
Trust	*t*(75)	0.23	0.80	−1.38	−1.94	0.68	−2.37	−2.23	−2.73	−3.19	−0.20
	p	0.821	0.428	0.171	0.056	0.498	0.020	0.029	0.008	0.002	0.846
	Cohens *d*	0.03	0.09	0.16	0.22	0.08	0.27	0.26	0.31	0.37	0.02
Satisfaction	*t*(75)	1.01	2.58	−1.77	−1.39	2.22	−3.82	−2.90	−5.02	−4.40	0.51
	p	0.315	0.012	0.081	0.168	0.030	0.000	0.005	0.000	0.000	0.614
	Cohens *d*	0.12	0.30	0.20	0.16	0.25	0.44	0.33	0.58	0.50	0.06
Continue in system	*t*(75)	1.01	2.13	−1.83	−1.19	1.71	−4.03	−2.92	−5.06	−3.93	0.76
	p	0.318	0.036	0.072	0.237	0.091	0.000	0.005	0.000	0.000	0.449
	Cohens *d*	0.12	0.24	0.21	0.14	0.20	0.46	0.33	0.58	0.45	0.09

Note: Two-tailed paired *t*-tests
NP system without punishment, *PPIC* peer punishment with individual costs, *PPSC* peer punishment with shared costs, *DPIC* democratic punishment with individual costs, *DPSC* democratic punishment with shared costs

Perceived procedural justice of the reduction process (i.e., Procedural Justice 2) was significantly lower in the peer punishment system compared to the democratic punishment system.

The democratic punishment system and the system without punishment did not differ significantly from each other and resulted in the highest satisfaction rates. Satisfaction in the peer punishment system was significantly lower compared to the democratic system and the system without punishment.

Trust in the democratic punishment system did not differ significantly from the system without punishment, but was significantly higher compared to the peer punishment system. Trust between the system without punishment and the peer punishment system did not significantly differ. Finally, considering the willingness to continue in a system, the democratic system and the system without punishment options had the highest ratings and differed significantly from the peer punishment system.

In summary, the results replicate previous findings and support the assumption that the democratic peer punishment systems inhibit the execution of antisocial punishment, and at the same time establish high levels of cooperation. Moreover, they extend previous findings showing that the systems can be distinguished by their specific relation to psychological constructs, especially to distributive and procedural justice.

Introducing a peer punishment system with shared costs (PPSC; H5–H7)

Contribution. Average cooperation rates in the PPSC was significantly higher compared to the system without punishment, but were not significantly different from the peer punishment system and the democratic punishment system.

Punishment. Considering separately the intended altruistic and antisocial punishment, there was a significant difference between the systems of intended antisocial punishment. In line with our hypothesis, this was driven by more antisocial punishment directed towards cooperative individuals in the PPSC (H5). There was also a significant difference in the executed punishment. In line with our hypothesis, this was again driven by more antisocial punishment directed towards cooperative individuals in both peer punishment system (H5).

Total payoff. Again, in the PPSC the higher cooperation levels did not cover the costs of punishment. Contrary to our assumption,

the total payoff in the PPSC did not differ significantly from the peer punishment system, but was significantly lower compared to the democratic punishment system and the system without punishment (H6), which aligned with our assumption.

Psychological constructs. Distributive justice was significantly lower in the PPSC compared to all other systems. Perceived participation possibility (i.e., Procedural Justice 1) was significantly higher compared to the system without punishment options but did not differ significantly from any other system with punishment. Perceived procedural justice of the reduction process (i.e., Procedural Justice 2) was significantly lower in the PPSC compared to the democratic punishment system but did not differ significantly from the peer punishment system.

Satisfaction and willingness to continue in the system was significantly lower in the PPSC compared to all other systems. Trust in the PPSC was significantly lower compared to the democratic punishment system and did not differ significantly from the system without punishment and from the peer punishment system.

In sum, the results indicate that a peer punishment system with shared costs can establish high levels of cooperation but facilitates the emergence of high antisocial punishment and thus resulting in low payoff. In addition, the system is experienced as highly unjust considering distributive justice and justice of the reduction process. Moreover, participants in this system experienced less trust and satisfaction and indicated the lowest willingness to keep on playing.

Introducing a democratic punishment system with shared costs (DPSC; H8–H10)

Contribution. Again, average cooperation rates in the DPSC was significantly higher compared to the system without punishment and the peer punishment system. There were no significant differences among the DPSC, the PPSC, and the democratic system.

Punishment. Considering separately the intended altruistic and antisocial punishment, there was a significant difference between the systems of intended, as well as executed antisocial punishment. Consistent with our assumptions, this was driven by more antisocial punishment directed towards cooperative individuals in punishment systems without voting. However, contrary to our hypothesis, there was no significant difference between the democratic system and the DPSC.

Total payoff. In the DPSC the total payoff was higher compared to the peer punishment system and PPSC. However, the payoff did not further increase compared to the system without punishment and the democratic system.

Psychological constructs. Distributive justice and justice of the reduction process (i.e., Procedural Justice 2) was significantly higher in the DPSC compared to all other systems, except for the democratic system. Perceived participation possibility (i.e., Procedural Justice 1) was significantly higher compared to the system without punishment options but did not significantly differ from any other system with punishment.

Satisfaction and willingness to continue was significantly higher in the DPSC compared to systems without voting but not significantly different from the system without sanctions and the democratic system. Trust in the DPSC was significantly higher compared to the systems without voting and did not differ significantly from the system without punishment and the democratic system.

In sum, the results indicate that a democratic punishment system with shared costs can establish high levels of cooperation, and at the same time prevent the execution of antisocial punishment and thus resulting in higher payoff. In addition, the system is experienced as just considering distributive justice, participation possibilities, and justice of the reduction process. Moreover, participants experienced more trust and satisfaction and indicated a high willingness to keep on playing in this system.

Willingness to continue (H11 and H12). Finally, considering the willingness to continue in a system, the PPSC had significantly lower ratings than all other systems. Both democratic systems and the system without punishment options had the highest ratings. Willingness to continue in the peer punishment system with individual costs was significantly higher than in the peer punishment system with shared costs.

In all five systems, willingness to continue was positively correlated with all psychological constructs, as well as income and own contribution. The only exception was a positive but not significant relation between participation possibilities and willingness to continue in the democratic punishment system. Additionally, in the system without punishment, trust was unrelated to the willingness to continue (all correlations are displayed in Tables 4 and 5).

Finally, we conducted multiple regressions using the psychological factors, as well as mean income in a system as predictors. The results revealed that neither monetary income nor distributive justice or trust but rather satisfaction is the strongest (and in some systems sole) predictor of willingness to continue in any system (Table 6). The implications of this finding for designing institutions to overcome social dilemmas will be discussed in the "Discussion" section.

Discussion

The present research focused on how the characteristics of different punishment systems (i.e., peer punishment versus democratic punishment and individual versus shared costs) influence cooperative behavior in a social dilemma. Moreover, the aim of the current study was to investigate potential psychological benefits of democratic punishment systems compared to peer punishment systems or systems without punishment. Therefore, we investigated not only the relationship between system characteristics and monetary outcomes, but also how these characteristics relate to justice perceptions, satisfaction, and trust, as well as the willingness to stay in a given system.

To test our assumptions, we compared five systems which differed in the possibility to punish other group members, in the punishment decision-making process, and in the distribution of the punishment costs. Participants' subjective experiences in each system were measured on several psychologically important constructs. Our results show that democratic systems compared to other systems promote cooperative behavior in social dilemmas. Replicating previous findings (Fehr and Gächter, 2002; Pfattheicher et al., 2018), our results indicate that in the long run, punishment pays off in the sense that it stabilizes and maintains high cooperation rates. Additionally, democratic punishment in particular pays off earlier than peer punishment. Like previous studies (Ambrus et al., 2017; Pfattheicher et al., 2018), our findings also show that through a voting process, destructive behavior (i.e., antisocial punishment) can be prevented. Interestingly, and contrary to some of our assumptions, shared costs of punishment did not have a strong influence on contribution or punishment behavior.

Table 4 Pearson's correlation between the ratings of psychological constructs in each system

	Distributive Justice	Procedural Justice 1	Procedural Justice 2	Satisfaction	Trust
System with punishment options					
Procedural Justice 1	−0.02	-	-	-	-
Procedural Justice 2	n.a.	n.a.	-	-	-
Satisfaction	0.50***	0.13	n.a.	-	-
Trust	0.24*	−0.14	n.a.	0.31**	-
Willingness to continue	0.31**	0.19^{+}	n.a.	0.52***	0.08
Peer punishment with individual costs					
Procedural Justice 1	0.45***	-	-	-	-
Procedural Justice 2	0.61***	0.38***	-		
Satisfaction	0.63***	0.45***	0.59***	-	-
Trust	0.48***	0.21^{+}	0.15	0.51**	-
Willingness to continue	0.45***	0.34**	0.46***	0.52***	0.40***
Peer punishment with shared costs					
Procedural Justice 1	0.47***	-	-	-	-
Procedural Justice 2	0.70***	0.52***	-		
Satisfaction	0.75***	0.60***	0.59***	-	-
Trust	0.63***	0.45***	0.33**	0.62***	-
Willingness to continue	0.64***	0.40***	0.70***	0.73***	0.43***
Democratic punishment with individual costs					
Procedural Justice 1	0.19	-	-	-	-
Procedural Justice 2	0.61***	0.21^{+}	-		
Satisfaction	0.74***	0.37**	0.51***	-	-
Trust	0.51***	0.14	0.44***	0.51***	-
Willingness to continue	0.54***	0.14	0.46***	0.66***	0.40***
Democratic punishment with shared costs					
Procedural Justice 1	0.54***	-	-	-	-
Procedural Justice 2	0.75***	0.46***	-		
Satisfaction	0.75***	0.46***	0.66***	-	-
Trust	0.56***	0.41***	0.50***	0.56***	-
Willingness to continue	0.55***	0.47***	0.62***	0.65***	0.33***

Note. All variables were measured on a scale from (1) *not at all true* to (2) *completely true*
$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.00$

Table 5 Pearson's correlation between individuals' contribution in each system with ratings of psychological constructs in the corresponding system

	Distributive Justice	Procedural Justice 1	Procedural Justice 2	Satisfaction	Trust	Willingness to continue	Contribution
Own contribution in the ...							
No punishment system	0.14	0.24*	n.a.	0.12	−0.22^{+}	0.25*	-
Peer punishment system individual costs	0.37***	0.40***	0.32***	0.43***	0.34**	0.42***	-
Peer punishment system shared costs	0.29*	0.25*	0.29*	0.27*	0.33**	0.29*	-
Democratic peer punishment system individual costs	0.34**	0.20^{+}	0.41***	0.39***	0.26*	0.40***	-
Democratic peer punishment system shared costs	0.43***	0.42***	0.39**	0.37***	0.48***	0.35**	-
Own income in the ...							
No punishment system	0.20^{+}	0.31**	n.a.	0.26*	−0.26*	0.32**	0.84***
Peer punishment system individual costs	0.39**	0.40***	0.46***	0.57***	0.26*	0.39**	0.70***
Peer punishment system shared costs	0.51***	0.56***	0.57***	0.67***	0.45***	0.58***	0.60***
Democratic peer punishment system individual costs	0.49***	0.32**	0.69***	0.53***	0.33**	0.48***	0.62***
Democratic peer punishment system shared costs	0.55***	0.47***	0.60**	0.50***	0.48***	0.37**	0.76***

Note. All variables were measured on a scale from (1) *not at all true* to (2) *completely true*
$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.00$

Table 6 Multiple regression analyses

Predictor	Willingness to continue in the system				
	F	R^2	*B*	*SE B*	*β*
System without punishment options	6.71***	0.27			
Distributive Justice			0.21	0.17	0.17
Procedural Justice 1			−0.03	0.09	−0.04
Satisfaction			0.35	0.20	0.26^{+}
Trust			0.20	0.15	0.16
Income			0.02	0.02	0.14
Peer punishment with individual costs	5.25***	0.25			
Distributive Justice			0.06	0.15	0.05
Procedural Justice 1			−0.05	0.13	−0.04
Procedural Justice 2			−0.03	0.13	−0.03
Satisfaction			0.63	0.18	0.45**
Trust			−0.04	0.15	−0.03
Income			0.11	0.06	0.20^{+}
Peer punishment with shared costs	22.85***	0.64			
Distributive Justice			−0.08	0.17	−0.06
Procedural Justice 1			−0.22	0.11	−0.20*
Procedural Justice 2			0.50	0.12	0.46***
Satisfaction			0.72	0.16	0.56***
Trust			0.02	0.13	0.56
Income			0.02	0.03	0.09
Democratic punishment with individual costs	10.64***	0.44			
Distributive Justice			0.03	0.19	0.02
Procedural Justice 1			−0.16	0.11	−0.14
Procedural Justice 2			0.07	0.15	0.06
Satisfaction			.84	0.22	0.55***
Trust			0.06	0.12	0.05
Income			0.05	0.04	0.17
Democratic punishment with shared costs	13.11***	0.49			
Distributive Justice			−0.15	0.21	−0.11
Procedural Justice 1			0.26	0.12	0.21*
Procedural Justice 2			0.51	0.17	0.41**
Satisfaction			0.84	0.23	0.49***
Trust			−0.17	0.15	−0.12
Income			−0.03	0.03	−0.10

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.00$

Without voting, as we expected, more antisocial punishment emerged. This is probably based on the lower individual costs of punishment. In the democratic systems we found no differences in the punishment behavior between the shared and individual cost version. One possible explanation is that we expected a further reduction of executed antisocial punishment compared to the democratic system with individual cost, and thus a higher net benefit. However, the total amount of executed antisocial punishment in the democratic system was already quite low, and therefore a further (significant) reduction was unlikely. Considering the monetary outcome (i.e., the total payoff), the democratic systems were still just as good as the system without punishment. This finding poses an interesting question: If the monetary payoff is not better, then why should we bother to implement complex punishment systems?

We argue that the answer to this question is to not focus on the *monetary* outcome, but on the *immaterial* outcome. Our results show that psychological constructs like justice perceptions, satisfaction, and trust were highest in the democratic systems. Therefore, in the long run the implementation of a democratic punishment system has three main advantages compared to other systems: (1) stabilization and maintenance of high cooperative behavior; (2) high monetary payoffs; and, (3) high beneficial psychological consequences. All other systems fall short of a least one of these points. We believe that the third aspect in particular has often been neglected in social dilemma research.

Why does this matter? Unlike strict economic theories or assumptions about humans as rational beings (Simon, 1955, 1959), research from different areas has consistently shown that humans do not solely thrive to maximize monetary benefits (Gintis, 2000; Henrich et al., 2010; Tversky and Kahneman, 1974).Therefore, it is no surprise that in our study we also find that participants are willing to engage in costly punishment, and that they are willing to contribute to a common good, even so they can achieve the best individual outcome by defecting. Also, a somehow counterintuitive behavior is that people prefer systems with punishment opportunities over systems without punishment opportunities, even though the payoff is less (Gürerk et al., 2006). As stated in the "Introduction", there are at least two possible explanations: (1) people do not consciously realize the lower monetary payoff, and (2) they hope to gain some other benefits from the implementation of punishment systems.

Our findings support our assumption that there are others benefits like higher perceived justice, higher satisfaction, and higher trust. Of course, we have no possibility of knowing if people consciously evaluate and prefer a system based on these factors. However, the importance of these aspects is best shown by considering what is probably the most important result of our study: Participants indicated the highest willingness to continue in the democratic systems, even though the monetary payoff did not significantly differ from the system without punishment opportunities. Additionally, in all systems it was not monetary outcomes but satisfaction that emerged as the best predictor for a willingness to stay in a given system. Satisfaction was positively related to justice perceptions and trust, but also to individual income and contribution, indicating that satisfaction is not completely independent of monetary outcomes.

Our findings indicate that the downfall of cooperative behavior is accompanied by a downfall in trust and satisfaction, which makes future interaction in that system unlikely. Maybe given the opportunity, people would not even choose to play in that system. Many studies have investigated so-called exit options in social dilemmas (Dana et al., 2006; Seale et al., 2006; Van Vugt and Hart, 2004). In these scenarios, participants are presented with a choice between allocating goods between themselves and others or leaving without the other players knowing. Results show that participants often prefer the exit option, even though they could take the same or more money by making a decision (Engel, 2011). These findings also indicate that there is more to satisfaction than monetary outcomes.

Furthermore, experienced satisfaction might be transferred to future interaction. The findings by Peysakhovich and Rand (2015) showed that in social dilemmas people use experiences in previous interactions as guidelines for behavior in future interactions. In combination with our findings it seems plausible that people are more willing to opt into systems in which they have previously had pleasant and beneficial experiences. Therefore, we believe a change of perspective is needed. Instead of focusing on enhancing monetary payoffs or proposing rewards for cooperative behavior (Andreoni and Gee, 2012; Ouss and Peysakhovich, 2015), studies should begin to focus more on psychological aspects that influence the willingness to keep on being cooperative in the future.

Limitations and outlook. We first want to mention methodological shortcomings of the present study. In this regard, we would like to acknowledge that we chose a within design to have higher statistical power, however, we had a small sample size. A between-subject design would have been preferable, but this would have led to a tremendous explosion of sample size. Moreover, to prevent carry-over effects, players' numbers were randomly changed per period, but players remained in the same group of four players. With this method we only controlled for possible carry-over effects on the individual level, but not on the group level. Although we did not find any order effects in our study, a between design would definitely prevent any carry-over effects between the conditions. To replicate our findings and strengthen our claims, a study with a between subject design and a larger sample size should be conducted.

In addition, our study was conducted in Germany, a representative democracy, in which each citizen is allowed to vote for representatives who are in turn supposed to promote their constituents' view in the parliament. The social heuristic hypothesis (Rand et al., 2014) assumes that people in a new context behave intuitively in a way that proves beneficial in their daily lives. Therefore, the familiarity with the system could also contribute to the satisfaction and the willingness to stay in this system. However, we measured participants' willingness to stay after ten periods. Thus, participants also had enough time to decide deliberately and to overcome heuristic thinking. It remains unclear if our findings could be replicated in other societies, especially in non-WEIRD societies (Henrich et al., 2010). Nevertheless, the overall finding that satisfaction is the best predictor for willingness to stay in a given system might be replicated and generalizable across societies.

Apart from satisfaction, the other psychological constructs did not predict willingness to continue in a system. We assume that justice perceptions are at least to some extent predecessors of satisfaction. However, our design does not allow drawing any causal inference between the constructs. Future studies could address this point.

In comparison to numerous other studies, our research does not take into account any individual differences. In their seminal paper about the origins of prosocial preferences, Rand et al. (2012) argue that people do not have a single set of social preferences, but rather that more complex models of cooperation are needed. In line with the argument, studies show that prosociality is not per se driven by social preferences, but also influenced by the consideration of moral righteousness of a behavior (Capraro and Rand, 2018; Eriksson et al., 2017) and social norms (Krupka and Weber, 2013). Like these authors, we also believe that cooperative behavior may be a result of complex interactions of various factors, and that there is no sole factor that determines cooperative behavior. Certainly there exist some individual factors which relate to a general tendency to be more or less cooperative, and there are also some contextual aspects which promote or prevent the emergence of cooperative behavior. On a theoretical level, we argue that social preferences (like SVO, for example) and their relation to cooperation should be similar across institutions (i.e., different punishment systems), meaning that a prosocial individual will probably on average cooperate in all systems more than will a pro-self individual. Applying the same logic, if intuitive decision making is related to cooperative behavior (Rand et al., 2014), then we would assume that in all systems, intuitive decision making leads to higher cooperation compared to deliberative decision making, although the absolute cooperation levels might differ between the systems. However, how the perception and the preference of a specific system relate to individual differences remains unclear. We leave this question to further research.

Our research aimed to detect one more *contextual* piece of the puzzle to answer the question why people cooperate. We argue and show that the characteristics of institutional designs (i.e., implementation and structure of punishment systems) influence cooperative behavior, but additionally these characteristics also relate to psychological constructs. By taking psychological benefits into account, cooperative behavior might be more understandable and easier to predict. In this way, our findings might help to understand why sometimes it appears to be irrational of people to cooperate considering monetary outcome. We want to emphasize that our research does not imply that monetary (or material) outcomes are unimportant for cooperation. We argue instead that monetary outcomes *alone* are not enough to explain individuals' preferences for (democratic) punishment systems and cooperative behavior in the long run. Therefore, our results can provide ideas for policy makers and help to design institutions which promote peoples' willingness to be cooperative.

Hauser et al. (2014) concluded that "Many citizens are ready to sacrifice for the greater good. We just need institutions that help them do so." Our empirical investigation focused on several psychological factors like satisfaction, which influence peoples' willingness to be cooperative and support for an institution. In sum, the current work provides new insights for understanding why people are willing to choose punishment systems and for how to design systems that facilitate cooperation; moreover, it offers new approaches to measure the beneficial effects of democratic punishment systems beyond solely monetary outcomes.

Received: 19 November 2018 Accepted: 3 April 2019
Published online: 07 May 2019

Notes

1 We did not implement all 120 possible orders, but rather we implemented five different orders to ensure that each system was played at each possible point (as the first, the second, the third, the fourth, or the fifth) during the experiment. We found no significant effect of order on any of the dependent variables.

References

Ambrus A, Greiner B, Sastro A (2017) The case for nil votes: Voter behavior under asymmetric information in compulsory and voluntary voting systems. J Public Econ 154:34–48. https://doi.org/10.1016/j.jpubeco.2017.08.006

Andreoni J, Gee LK (2012) Gun for hire: delegated enforcement and peer punishment in public goods provision. J Public Econ 96(11–12):1036–1046

Balliet D, Van Lange PA (2013a) Trust, conflict, and cooperation: a meta-analysis. Psychol Bull 139(5):1090–1112. https://doi.org/10.1037/a0030939

Balliet D, Van Lange PA (2013b) Trust, punishment, and cooperation across 18 societies: a meta-analysis. Perspect Psychol Sci 8(4):363–379. https://doi.org/10.1177/1745691613488533

Balliet D, Mulder LB, Van Lange PA (2011) Reward, punishment, and cooperation: a meta-analysis. Psychol Bull 137(4):594. https://doi.org/10.1037/a0023489

Balliet D, Parks C, Joireman J (2009) Social value orientation and cooperation in social dilemmas: a meta-analysis. Group Process Inter Relat 12(4):533–547

Barber B 1983: The Logic and Limits of Trust. New Brunswick: Rutgers University Press

Barcelo H, Capraro V (2015) Group size effect on cooperation in one-shot social dilemmas. Sci Rep 5:7937

Beierlein C, Kemper C, Kovaleva A, Rammstedt B (2012). Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3). Köln: GESIS

Bies RJ, Martin CL, Brockner J (1993) Just laid off, but still a "good citizen?" Only if the process is fair. Empl Responsib Rights J 6(3):227–238

Bogaert S, Boone C, Declerck C (2008) Social value orientation and cooperation in social dilemmas: a review and conceptual model. Br J Soc Psychol 47 (3):453–480

Brockner J, Grover S, Reed T, DeWitt R, O'Malley M (1987). Survivors' reactions to layoffs: we get by with a little help for our friends. Adm Sci Q (4) 526–541

Camerer CF (2003) Behavioral game theory: experiments in strategic interaction. Russell Sage Foundation, New York, NY

Capraro V, Rand DG (2018). Do the right thing: experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. Judgement and Decision Making, 13 (1). pp. 99-111
Capraro V, Smyth C, Mylona K, Niblo GA (2014) Benevolent characteristics promote cooperative behaviour among humans. PLoS One 9(8):e102881
Cohen J (1992) A power primer. Psychol Bull 112(1):155
Cohen-Charash Y, Spector PE (2001) The role of justice in organizations: a meta-analysis. Organ Behav Hum Decis Process 86(2):278–321
Colquitt JA, Scott BA, Judge TA, Shaw JC (2006) Justice and personality: using integrative theories to derive moderators of justice effects. Organ Behav Hum Decis Process 100(1):110–127
Dana J, Cain DM, Dawes RM (2006) What you don't know won't hurt me: costly (but quiet) exit in dictator games. Organ Behav Hum Decis Process 100 (2):193–201
Dasgupta P (1988) Trust as a commodity. In: Gambetta D (Ed.) Trust: making and breaking cooperative relations. Basil Blackwell, New York, NY, p 47–72
Dawes RM (1980) Social dilemmas. Annu Rev Psychol 31(1):169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125
Deutsch M (1960) The effect of motivational orientation upon trust and suspicion. Hum Relat 13(2):123–139
Dreber A, Rand DG, Fudenberg D, Nowak MA (2008) Winners don't punish. Nature 452:348–351. https://doi.org/10.1038/nature06723
Egas M, Riedl A (2008). The economics of altruistic punishment and the maintenance of cooperation. Proc R Soc B 275(1637), 871–878. https://doi.org/10.1098/rspb.2007.1558
Engel C (2011) Dictator games: a meta study. Exp Econ 14(4):583–610. https://doi.org/10.1007/s10683-011-9283-7
Eriksson K, Strimling P, Andersson PA, Lindholm T (2017) Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction. Journal of Experimental Social Psychology 69:59–64
Ertan A, Page T, Putterman L (2009) Who to punish? Individual decisions and majority rule in mitigating the free rider problem. Eur Econ Rev 53 (5):495–511. https://doi.org/10.1016/j.euroecorev.2008.09.007
Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res methods 41(4):1149–1160
Fehr E, Gächter S (2002) Altruistic punishment in humans. Nature 415:137–140. https://doi.org/10.1038/415137a
Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. The quarterly journal of economics 114(3):817–868
Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10(2):171–178
Fitzsimons GJ, Greenleaf EA, Lehmann DR (1997). Decision and consumption satisfaction: Implications for channel relations. Marketing Studies Center Working Paper Series, 313.
Folger R, Konovsky MA (1989) Effects of procedural and distributive justice on reactions to pay raise decisions. Acad Manag J 32(1):115–130
Gächter S, Renner E, Sefton M (2008) The long-run benefits of punishment. Science 322:1510–1510. https://doi.org/10.1126/science.1164744
Gintis H (2000) Strong reciprocity and human sociality. J Theor Biol 206 (2):169–179
Guala F (2012) Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. Behav Brain Sci 35(1):1–15. https://doi.org/10.1017/S0140525X11000069
Gürerk Ö, Irlenbusch B, Rockenbach B (2006) The competitive advantage of sanctioning institutions. Science 312:108–111
Hardin G (1968) The tragedy of the commons. Science 162:1243–1248. https://doi.org/10.1126/science.162.3859.1243. 1968
Hauser OP, Rand DG, Peysakhovich A, Nowak MA (2014) Cooperating with the future. Nature 511(7508):220
Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behav Brain Sci 33(2–3):61–83
Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. Science 319(5868):1362–1367. https://doi.org/10.1126/science.1153808
Hilbe C, Traulsen A, Röhl T, Milinski M (2014). Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. Proc Natl Acad Sci USA 111(2), 752–756. https://doi.org/10.1073/pnas.1315273111.
Hilbig BE, Zettler I, Heydasch T (2012) Personality, punishment and public goods: strategic shifts towards cooperation as a matter of dispositional honesty–humility. Eur J Personal 26(3):245–254
Kim O, Walker M (1984) The free rider problem: experimental evidence. Public Choice 43(1):3–24. https://doi.org/10.1007/BF00137902
Krupka EL, Weber RA (2013) Identifying social norms using coordination games: Why does dictator game sharing vary? Journal of the European Economic Association 11(3):495–524
McAllister DJ (1995) Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 38(1):24–59
Messick DM, Sentis K (1983) Fairness, preference, and fairness biases.In: David M. Messick and Karen S. Cook (eds.), Equity theory: psychological and sociological perspectives. p. 61–94. New York: Praeger
Moorman RH (1991) Relationship between organizational justice and organizational citizenship behaviors: do fairness perceptions influence employee citizenship? J Appl Psychol 76(6):845
Mulder LB, Van Dijk E, De Cremer D, Wilke HA (2006) Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. Journal of Experimental social psychology 42(2):147–162
Nikiforakis N, Normann HT (2008) A comparative statics analysis of punishment in public-good experiments. Exp Econ 11(4):358–369. https://doi.org/10.1007/s10683-007-9171-3
Ostrom E (1990) Governing the commons: the evolution of institutions for collective action. Cambridge University Press, New York, NY
Ouss A, Peysakhovich A (2015) When punishment doesn't pay: cold glow and decisions to punish. J Law Econ 58(3):625–655
Peysakhovich A, Rand DG (2015) Habits of virtue: creating norms of cooperation and defection in the laboratory. Manag Sci 62(3):631–647
Pfattheicher S, Böhm R, Kesberg R (2018) The advantage of democratic peer punishment in sustaining cooperation within groups. J Behav Decis Mak 31 (4):562–571. https://doi.org/10.1002/bdm.2050
Pillai R, Williams ES, Justin Tan J (2001) Are the scales tipped in favor of procedural or distributive justice? An investigation of the US, India, Germany, and Hong Kong (China). Int J Confl Manag 12(4):312–332
Price JL, Mueller CW (1986) Absenteeism and turnover of hospital employees. JAI, Greenwich, CT
Rabin M (1993). Incorporating fairness into game theory and economics. The American economic review, 1281–1302
Rand DG, Nowak MA (2011) The evolution of antisocial punishment in optional public goods games. Nature communications 2:434
Rand DG, Peysakhovich A., Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, Greene JD (2014) Social heuristics shape intuitive cooperation. Nature communications 5:3677
Rand DG, Nowak MA (2013) Human cooperation. Trends Cogn Sci 17 (8):413–425. https://doi.org/10.1016/j.tics.2013.06.003
Rand DG, Greene JD, Nowak MA (2012) Spontaneous giving and calculated greed. Nature 489(7416):427
Schurter K, Wilson BJ (2009) Justice and fairness in the dictator game. South Econ J 76(1):130–145
Seale DA, Arend RJ, Phelan S (2006) Modeling alliance activity: opportunity cost effects and manipulations in an iterated prisoner's dilemma with exit option. Organ Behav Hum Decis Process 100(1):60–75
Simon HA (1955) A behavioral model of rational choice. Q J Econ 69(1):99–118
Simon HA (1959) Theories of decision-making in economics and behavioral science. Am Econ Rev 49(3):253–283
Sitkin SB, Roth NL (1993) Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. Organ Sci 4(3):367–392
Simpson JA (2007) Psychological foundations of trust. Curr Dir Psychol Sci 16 (5):264–268
Tinghög G, Andersson D, Bonn C, Böttiger H, Josephson C, Lundgren G, Johannesson M (2013) Intuition and cooperation reconsidered. Nature 498 (7452):E1
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Van den Bos K (1999) What are we talking about when we talk about no-voice procedures? On the psychology of the fair outcome effect. J Exp Soc Psychol 35(6):560–577
Van den Bos K, Lind EA (2002) Uncertainty management by means of fairness judgments. In: M. P. Zanna (Ed.), Advances in experimental social psychology, vol. 34. pp. 1–60. New York Academic Press
Van den Bos K, Lind EA, Vermunt R, Wilke HA (1997) How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. J Personal Soc Psychol 72(5):1034
Van den Bos K, Vermunt R, Wilke HA (1997) Procedural and distributive justice: what is fair depends more on what comes first than on what comes next. J Personal Soc Psychol 72(1):95
Van Lange PA, Joireman J, Parks CD, Van Dijk E (2013) The psychology of social dilemmas: a review. Organ Behav Hum Decis Process 120(2):125–141. https://doi.org/10.1016/j.obhdp.2012.11.003
Van Lange PA, Vugt MV, Meertens RM, Ruiter RA (1998) A social dilemma analysis of commuting preferences: the roles of social value orientation and trust. J Appl Soc Psychol 28(9):796–820
Van Vugt M, Hart CM (2004) Social identity as social glue: the origins of group loyalty. J Personal Soc Psychol 86(4):585
Yamagishi T (2011). Trust: the evolutionary game of mind and society New York, NY: Springer

Zucker LG (1986) Production of trust: institutional sources of economic structure, 1840-1920. Res Organ Behav 8:53–111

Acknowledgements
This research was supported by a grant from the Baden-Württemberg Foundation to the second author.

Additional information
Supplementary information: The online version of this article (https://doi.org/10.1057/s41599-019-0249-2) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare that the research was conducted in the abscence of any commercial or financial relationships that could be construied as a potential conflict of interest.

Reprints and permission information is available online at http://www.nature.com/reprints

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.