Cooperation and social control

Bakker, Dieko Marnix

DOI:
10.33612/diss.98552819

5

Multiple group membership in public good problems

Dieko M. Bakker, Jacob Dijkstra, Andreas Flache

This chapter is currently under review at an international peer-reviewed journal

Human societies cannot thrive without the provision of goods that require cooperation between individuals. These collective goods benefit all individuals involved, but cooperation is costly for those who contribute and cannot easily be enforced (Olson, 1965). 'Free rider' problems in the workplace (Petersen, 1992) and environmental pollution are two examples where cooperation fails due to the conflict between individual and collective interests (Bouma et al., 2008). This tension between individual and collective interests is a core property of social dilemmas (Dawes, 1980; Kollock, 1998; Van Lange et al., 2013). In social dilemmas, individually rational behavior often leads to collectively suboptimal outcomes. Social dilemmas contain opportunities for free-riders to take advantage of cooperative group members (Dawes, 1980; Fehr & Gächter, 2002; Kollock, 1998). Yet, in many social dilemmas a degree of cooperation is observed that is hard to explain by self-interest alone (Bowles et al., 2012), as exemplified by zealous cooperation in 'self-managing' work-teams (Barker, 1993) or successful management of common natural resources by local communities (Ostrom, Burger, Fields, Norgaard, & Policansky, 1999).

Social dilemmas generally occur within a social group. Research on social dilemmas thus focuses on processes within this group which influence whether the dilemma is successfully solved. Over the years, an extensive literature has developed around a joint interest in explaining and encouraging cooperation in social dilemmas. Much of the literature on social dilemmas has been focused on how to discourage free-rider behavior and promote behavior which furthers the collective interest (Balliet et al., 2011; Chaudhuri, 2011; Van Lange, Balliet, et al., 2014; Van Lange, Rockenbach, et al., 2014). Several solutions to this problem have been suggested, such as introducing communication between those involved in the dilemma (Balliet, 2010; Kollock, 1998), embedding the social dilemma in a recurring interaction (Axelrod, 1984; Friedman, 1971; Kollock, 1998) and creating reputation mechanisms (Diekmann et al., 2014; Milinski, Semmann, & Krambeck, 2002) or selective incentives (Olson, 1965). The most common way to create such selective incentives is through reward or punishment (Balliet et al., 2011; Fehr & Gächter, 2002; Oliver, 1980; Van Lange, Rockenbach, et al., 2014). In general there is a prevailing view that these punishments are distributed by strong reciprocators: individuals who have a preference for mutual cooperation and are willing to bear the costs of enforcing cooperation in others (Bowles & Gintis, 2011; Fehr, Fischbacher, & Gächter, 2002; Gintis, 2000; Simpson & Willer, 2015).

Our collective understanding of behavior in social dilemmas, particularly in standardized situations with small groups, is now quite substantial. The next step is to

explore the generalizability of this knowledge. A movement in this direction has already begun. There is an ongoing discussion about the applicability of knowledge gained from laboratory experiments to real-life groups (e.g. Guala, 2012 and responses). The limits of punishment (in the presence of retaliation, e.g. Nikiforakis, 2008; Nikiforakis & Engelmann, 2011) and reward (in the presence of reward exchange, e.g. Flache, Bakker, Mäs, & Dijkstra, 2017; Flache & Macy, 1996) as enforcers of cooperation are being explored. There are attempts to apply the knowledge gained from laboratory experiments to increasingly practical situations (e.g. Englmaier & Gebhardt, 2016; Fehr & Leibbrandt, 2011; Kraft-Todd, Yoeli, Bhanot, & Rand, 2015).

We believe that there is an additional aspect to the generalizability of research on social dilemmas, which thus far has received little attention. This aspect is the broader social environment within which a group is embedded, specifically the possibility of overlap in group membership. Individuals often belong to multiple groups with conflicting collective interests. For persons with cooperative intentions, this creates ambiguity: it is unclear to which collective good they should contribute and which behavior they should reward or punish. Should a firefighter in a disaster-stricken town report to the fire station and come to the aid of the town as a whole, or should she attempt to protect her own family first (e.g. Killian, 1952)? Others face similar ambiguity when deciding whether their fellow group members who display apparently uncooperative behavior should be reprimanded. Is an employee who leaves early from a team meeting a contemptable shirker, or does he leave early to volunteer at the local food bank? This distinction can be difficult to make, particularly since it is often impossible to perfectly monitor the behavior of group members outside of the shared group context.

Most of the existing literature on social dilemmas is not directly applicable to such situations, because it relies on a simple interdependency structure: individuals belonging to a single group with a single collective good. By ignoring the broader social structure surrounding the group, this literature in effect assumes that people decide about their behavior in each social dilemma in isolation, as if an employee arrives at work and entirely forgets that they have a family until their shift is over. Under the more complex interdependency structure described in the examples above, the behavior of a person within one group cannot be understood without taking into account their involvement in other social groups. In the present study we will investigate how overlap in group membership affects cooperation and sanctioning

behavior in social dilemmas, as a result of groups' inability to effectively monitor their members' behavior in other groups.

The present study makes several contributions to the literature. First, to our knowledge, this is the first experimental investigation of behavior in multiple simultaneous public good games with overlap in group membership. Second, we introduce and validate a punishment system suitable for such an experiment. Third, we investigate how imperfect monitoring of free-riders caused by multiple group membership impacts sanctioning and contribution behavior.

Overview of the literature on overlap in group membership

Because the topic of multiple group membership is relatively new to the social dilemma literature, we first want to illustrate the relevance of overlap in group membership to central topics in social dilemma research. Such topics include cooperation, intergroup conflict, and social control. Before investigating the main research question of this study, we will therefore give an overview of literature outside of the field of social dilemmas which has demonstrated the importance of multiple group membership. Then, we will note some relevant aspects of the social dilemma literature which can serve as starting points for an investigation into the impact of the broader social structure on social dilemma behavior. There is an extensive literature on the effects of multiple group membership outside of research on social dilemmas. By contrast, there has been surprisingly little consideration of the importance of this topic in social dilemma research. This contrast illustrates a gap in the social dilemma literature in which there is a lot of potential for advancement.

Social psychology: social identity and intergroup bias

The relevance of (multiple) group membership to social behavior has been demonstrated in numerous fields of social science. For instance, according to Social Identity Theory, group identification is an important determinant of intergroup conflict and overlap in membership has been suggested as a way to reduce intergroup discrimination (Tajfel, 1982; Tajfel & Turner, 1979). In the context of a common resource, which is shared by members of an overarching group with several subgroups, promoting the salience of the overarching group results in more restraint in the use of the common resource (Kollock, 1998; Kramer & Brewer, 1984). Similarly, intergroup bias can be reduced by focusing on shared elements of identity, such as common interests, rather than on elements of identities which divide a group into clear subgroups (Gaertner, Mann, Murrell, & Dovidio, 1989; Stark & Flache, 2012).

Organizational studies: conflicting demands, team and employee outcomes

Organizational scholars have identified the challenges and opportunities generated by overlapping group membership in organizations (Ashforth & Mael, 1989; O'Leary, Woolley, et al., 2011; Williams, 2001). Because individuals are often members of many different groups, they tend to experience inconsistent demands and competing identities (Ashforth & Mael, 1989). At the same time, within organizations, individuals who are part of multiple teams can facilitate cross-team information sharing and coordination (Milliken & Martins, 1996; O'Leary, Woolley, et al., 2011). Multiple team membership can improve performance through information sharing, increased attention to efficient work practices, and more opportunities for innovation (Bertolotti, Mattarelli, Vignoli, & Macrì, 2015; O'Leary, Woolley, et al., 2011). At the same time, multiple team membership poses unique challenges by requiring employees to deal with more frequent interruptions, more task switching, and a need to balance their available time and resources between multiple teams (Bertolotti et al., 2015). Employees may need time to adjust to multiple team membership before being able to take advantage of its opportunities (van de Brake, Walter, Rink, Essens, & van der Vegt, 2018) and performance may be negatively impacted if employees are asked to participate in too many teams at once (Bertolotti et al., 2015; O'Leary, Mortensen, & Woolley, 2011; O'Leary et al., 2011).

In the organizational literature, the term 'faultlines' has been used to describe how groups can divide into subgroups along lines determined by demographic characteristics (Lau & Murnighan, 1998). Such faultlines can lead to intragroup conflict and hence to decreased team performance (Lau & Murnighan, 2005). In such a situation, individuals who share characteristics with several subgroups bridge faultlines and may reduce conflict between subgroups (Lau & Murnighan, 2005; Mäs et al., 2013).

Sociology: Simmel, Blau & Merton

In Georg Simmel's account of the development of individual personality and social structure, the development of society is characterized by a shift from concentric circles to intersecting circles (Diani, 2000; Simmel, 1908; Stark & Flache, 2012). Both types of 'circles' or groups imply multiple group memberships. When concentric circles are predominant, multiple memberships are nested, as when family relationships and work relationships all take place within a closed local community. When intersecting circles are predominant, multiple memberships are more diverse and individualized, creating conflict between competing affiliations and loyalties (Diani, 2000; Simmel,

1908; Stark & Flache, 2012). The literature on social movements has taken up the importance of overlap in group memberships quite strongly, especially in regard to networks of movements (Diani, 2000). For example, competing pressures from multiple social groups are relevant to investigations of the effect of social ties on movement membership and activism (McAdam & Paulsen, 1993).

While Simmel suggests that intergroup conflict is more likely when group memberships are more individualized and less consistent, Peter Blau's theory of cross-cutting social circles (Blau, 1977; Blau & Schwartz, 1984) points to potential advantages of diverse multiple memberships. Blau emphasizes that every person belongs to multiple potentially overlapping social categories. When these social categories are mostly uncorrelated, group boundaries are not clearly defined. Intergroup relations are then common and shared characteristics form the basis for integration between social groups (Blau, 1977; Blau & Schwartz, 1984). However, when social categories are strongly correlated this reinforces group boundaries and inhibits intergroup contact and integration (Blau, 1977; Blau & Schwartz, 1984; Stark & Flache, 2012).

There are also similarities to Merton's (1957) role sets and later work on role conflicts (e.g. Adler & Adler, 1987; House & Rizzo, 1972; Van Sell, Brief, & Schuler, 1981). Although Merton (1957) himself noted that role sets refer to multiple roles associated with a single position within a social group, the conflicting demands experienced by a person with multiple roles in a group are very similar to those experienced by a person with positions in multiple social groups.

Relevant research on social dilemmas

In the light of the dearth of attention for overlap in group membership in recent research on social dilemmas, it is interesting to note that some early publications did recognize its importance. A 1952 paper in the American Journal of Sociology shows the relevance of overlap in group membership to cooperative behavior (Killian, 1952). The study describes competing demands experienced by members of a disaster-stricken community. For example, local firefighters and police officers were torn between their obligations to the community as a whole and obligations to their own families. One firefighter remarks: *"All the rest of the firemen had relatives that were hurt, and they stayed with them. Naturally they looked after them. If it hadn't been that my wife was all right, this town probably would have burned up. It's hard to say, but I kind of believe I would have been looking after my family, too."* (Killian, 1952). A few years before, the same journal had published an experimental study which investigated the behavior of

individuals faced with competing demands from different social groups (Stouffer, 1949). This study proposes that every social group has a range of acceptable and prohibited behaviors, and individuals who are a member of multiple groups will try to exhibit behavior which is acceptable to all groups.

In an explanation of the successful production of public goods which focuses on individuals' dependence on the group and the group's control capacity, Hechter (1987) briefly touches on overlap in group membership. According to Hechter's theory, when a person is a member of multiple groups with competing interests, the person is expected to conform to the group with greater control capacity. When the control capacity of the groups is equal, the person will comply with obligations from the group on which they are most dependent.

The topic seems never to have gained real traction in social dilemma research, though, and those recent studies in which there are multiple groups at all do not consider overlap in membership. The possibility of *changes* in group membership has been investigated, for example by allowing participants in an experiment to choose between sanctioning institutions (Gurerk, Gürerk, Irlenbusch, & Rockenbach, 2006), by repeatedly sorting participants into different groups according to their behavior in the experiment (Chaudhuri, 2011; Gunnthorsdottir, Houser, & McCabe, 2007) or by allowing participants to endogenously sort themselves into groups (Charness & Yang, 2014; Chaudhuri, 2011; Page, Putterman, & Unel, 2005). However, none of these studies considered simultaneous membership in multiple groups.

The other main field of research in which multiple groups interact in a social dilemma is research on intergroup conflict. There is, generally, significant competition between groups placed in situations of intergroup conflict (Gunnthorsdottir & Rapoport, 2006; Mäs & Dijkstra, 2014). Competition is fierce even though group members do not necessarily harbor ill will towards members of another group (Mäs & Dijkstra, 2014). There is a small number of studies which illustrate how relevant overlap in group membership may be to research on intergroup competition. The presence of intergroup ties, which could be achieved by an overlap in membership, may decrease the extent of intergroup competition and discrimination (Mäs et al., 2013; Nelson, 1989; Takács, 2001). These studies, however, differ from the present study in that they focus on situations in which individuals are members of only one group but have ties to members of another group.

THEORY

The many fields of research in which overlapping group membership has been investigated offer viable starting points for theory-building on this topic in the context of social dilemmas. From social psychology, we could develop expectations about how the salience of the group identities affects contributions (Kramer & Brewer, 1984). From sociology we can apply ideas about the ways overlap in membership may reduce conflict between groups (Diani, 2000; Lau & Murnighan, 2005; Merton, 1957; Takács, 2001) but create conflict within groups (Rauhut & Winter, 2017), and about the properties of a group that influence which group wins out when the interests of multiple overlapping groups conflict (Hechter, 1987). But before we embark on these endeavors, we should consider how existing strands of social dilemma theory apply to situations with overlap in group membership. Only when existing theories cannot provide clear and accurate predictions of individual behavior and collective outcomes in social dilemmas with overlap in group membership do we need to develop a new theory. In the present study, we show that findings from studies on single groups cannot easily be generalized to a situation with overlap in group membership, by investigating the monitoring of free-riders in a multiple-group situation.

Multiple group membership hinders free-rider detection

We know that when given the opportunity to punish, a substantial proportion of individuals do indeed punish those group members who display uncooperative behavior even when distributing punishment is costly to the punisher (Bowles et al., 2012; Chaudhuri, 2011; Fehr & Gächter, 2002). This punishment effectively incentivizes cooperative behavior in social dilemmas (Balliet et al., 2011; Chaudhuri, 2011; Van Lange et al., 2013), at least in behavioral experiments if not always in real life (Guala, 2012). Fehr & Gächter (2002) suggest that punishment is emotionally motivated, triggered by anger towards free-riding group members. The punisher's beliefs about the intentions behind other group members' actions also matter (Charness & Rabin, 2002; Falk et al., 2008; Rabin, 1993). In general, punishments are often assumed to be distributed by strong reciprocators who value cooperation and want others to behave cooperatively as well (Bowles & Gintis, 2011; Fehr et al., 2002; Gintis, 2000; Simpson & Willer, 2015).

Effective punishment requires not just the presence of motivated punishers and the availability of a sanctioning mechanism, but also accurate monitoring of the behavior of other group members (Hechter, 1987). Imperfect monitoring makes it difficult to accurately identify non-cooperative group members, thus hindering the

effective distribution of punishment (Fischer et al., 2016; Grechenig et al., 2010; Van Miltenburg et al., 2017). The contributions of other group members can easily be misperceived, as when a person's seemingly small contribution to a joint project required a lot of work behind the scenes which is not observed by others (Van Miltenburg et al., 2017). Potential punishers who are concerned about punishing group members unfairly may refrain from punishing when they cannot accurately judge others' contributions and intentions, although results are mixed in this regard (Bornstein & Weisel, 2010; Patel et al., 2010; Van Miltenburg et al., 2017).

Similarly inaccurate perceptions can occur in the context of multiple groups with overlap in membership. Information about your fellow group members' behavior in other groups can be difficult to obtain. This makes it difficult to assess whether group members who exhibit non-cooperative behavior in a particular group do so because they are generally uncooperative or because they were constrained by another group. Patel et al. (2010) offer the example of staff not showing up to a meeting, which may be either because of conflicting obligations elsewhere or because they did not want to go, although they do not include actual conflicting obligations in their study. Membership of multiple groups thus offers free riders an opportunity to disguise their defection as cooperative behavior towards another group. For other group members, it is then difficult to determine the intentions behind uncooperative behavior. This uncertain relationship between a group member's intentions and their displayed behavior may lead to a breakdown of cooperation (Bendor & Mookherjee, 1987; Kollock, 1993). Given that punishments are considered to be motivated by anger towards uncooperative individuals (Fehr & Gächter, 2002; Gintis, 2000), and that the punisher's judgement of the other's intentions matters (Falk et al., 2008), an inability to judge whether defection reflects uncooperative intentions should hinder the distribution of punishment. In fact, we expect that potential punishers are likely to withhold punishment when they cannot judge accurately whether non-cooperative behavior reflects non-cooperative intentions.

When multiple groups exist, and there is overlap in membership between these groups, judging intentions from behavior is still possible as long as all parties know that the groups do not place mutually exclusive demands on their members. When work is light and there are no deadlines looming, employees will not have trouble devoting enough time to both their work and their family. Anyone shirking at work or coming home late can fairly be judged based on their actions. However, when the workload is heavy or when a family member falls ill and requires care, satisfying demands from the employer and family simultaneously becomes impossible. When group members do

not have the necessary resources to contribute fully to all groups they are a member of, both cooperators and non-cooperators will show free-rider-like behavior. Judging whether non-cooperative behavior reflects non-cooperative intentions becomes difficult, and potential punishers are likely to withhold punishment as a result.

Hypothesis 1: Participants with low contributions are less likely to be punished when participants cannot contribute fully to all groups in which they participate (under the assumption that this restriction is common knowledge)

From the literature on social dilemmas we know that when free riders are less likely to be punished, free riding is more common. For example, when punishment is discouraged by the presence of retaliation this may negatively impact contribution levels in a Public Good Game (Nikiforakis, 2008; Chapters 3 and 4 of this dissertation). Applying the same logic to situations with overlap in group membership, we expect more free riding when punishers cannot judge accurately whether non-cooperative behavior reflects non-cooperative intentions.

Hypothesis 2: Contributions will be lower when participants cannot contribute fully to all groups in which they participate

These hypotheses represent expectations based on findings from single-group social dilemmas, applied to multi-group social dilemmas with overlap in membership. They form the starting point of our investigation into the effects of overlap in membership on behavior in and outcomes of social dilemmas.

METHOD

Treatments

We test these hypotheses experimentally, using three different treatments. Two of these treatments implement multiple group membership, by making each participant a member of two groups. Table 5.1 gives an overview of the main features of each treatment. Each treatment consists of 10 periods in which participants play a variant of the public good game (Ledyard, 1995). The one-group treatment is included as a baseline to demonstrate the effectiveness of the punishment system we employ, which deviates from common sanctioning systems in public good games, to accommodate multiple group membership. This is discussed in more detail below. The crucial difference between the two-group treatments is in the number of points which

can be contributed to the public good in each group. In the two-group treatment with sufficient endowments, 20 points can be contributed to the public good in each group. The endowment in this treatment, 40 points, is enough to allow participants to contribute the maximum amount to both of the public good games they play. Thus, although participants in the two-group treatment with sufficient endowments play the public good game in two separate groups, the two public good games play as two completely separate games. Potential punishers know that if one of their fellow group members contributes little to the public good in their group, this cannot be because this group member was constrained by (their behavior in) the other group.

Table 5.1. Overview of experimental treatments

Treatment	*Number of groups*	*Endowment*	*Maximum contribution per group*
One-Group Baseline	1	20	20
Two-Group Sufficient Endowments	2	40	20
Two-Group Insufficient Endowments	2	40	40

In the two-group treatment with insufficient endowments, however, the maximum contribution to each public good game is 40 points. That is, every participant could contribute their whole endowment in one of the two groups. Under these circumstances, when a participant contributes less than the maximum amount to a particular group, other group members cannot tell whether the low contributor kept the points to themselves or contributed them to the public good in the other group. This creates the conditions under which an accurate assessment of the intentions behind others' uncooperative behavior is difficult, and under which we hypothesize that low contributions will be less likely to be punished and free riders are likely to take advantage of the situation. By comparing behavior in the two two-group treatments, we are able to test the two hypotheses regarding contribution levels and punishment behavior.

The group composition

In the one-group baseline treatment, participants were randomly divided into groups of 4 participants. Participants played the public good game for 10 rounds in either a partner treatment (in two pilot sessions) or a stranger treatment (in our main experiment), whereby at the start of each round participants are randomly regrouped. In the two-group treatments, participants were first randomly divided into their first group of 4 (the 'A' groups) and then deterministically placed into their second group of 4 (the 'B' groups). B-groups were composed such that no two group members in a B-group also shared an A-group. This is visualized in Figure 5.1. Participants were made aware of this structure by emphasizing that every participant is in two groups between which they are the only common member. In each round of the experiment, participants played the PGG in both groups. Participants played these public good games for 10 rounds in a stranger treatment. In all treatments, including the one-group treatments, the matching structure and duration of the experiment were common knowledge. To make this group structure possible, each session of the experiment consisted of exactly 16 participants.

Figure 5.1. Visualization of group allocations

A1	**A1**	**A1**	**A1**	**A2**	**A2**	**A2**	**A2**	**A3**	**A3**	**A3**	**A3**	**A4**	**A4**	**A4**	**A4**
B1	**B2**	**B3**	**B4**	**B1**	**B2**	**B3**	**B4**	**B1**	**B2**	**B3**	**B4**	**B1**	**B2**	**B3**	**B4**

The contribution stage

The first part of each period, in all treatments, is the contribution stage. In this stage, all participants are endowed with 20 points in the one-group treatment or 40 points in the two-group treatments. Participants then decide, simultaneously and without communicating, how many of their points to contribute to a group project. In the two-group treatments, they do this first for their A-group and then for their B-group, but before seeing any of the contributions of other group members. They can contribute anywhere from 0 (keeping all points to themselves) to 20 or 40 whole points (contributing all they have). Each point contributed is multiplied by 3, so that the total benefit to the group of each point contributed is 3 points. The collective productivity is thus maximized when all group members contribute all of their points. The points produced in the group are then evenly divided across all 4 group members. This implies that any point a participant contributes to the group results in a return of 0.75 points to each group member, for a net return of -0.25 for the participant who made

the contribution. Because contributions to the public good are collectively beneficial at a cost to the individual, high contributions are the measure of cooperation in a public good game.

The punishment stage

After participants have decided on their contributions to each group, they move on to the punishment stage. In the punishment stage, participants see the contributions of each of their group members. The punishment stage, just like the contribution stage, is played first for A-groups and then for B-groups.

In all treatments, the punishment stage provides every participant an opportunity to sanction their fellow group members. Punishment occurs by a collective decision rule (Van Miltenburg et al., 2014). Participants must decide, for each of their fellow group members, whether they want to punish that group member or not. When at least two group members vote to punish the same person, that person is excluded from the benefits of the public good. That is, punished participants forfeit any income from the group. Any points they invested in the public good are lost to them, and they do not benefit from the contributions of others. When a participant is punished, all group members who voted to punish this person pay a fee of 5 points. The points which would have gone to the punished participant are not redistributed among the other group members, so as not to give group members an incentive to exclude others for their personal benefit.

There are two main reasons to select this method of punishment, one substantive and one design related. First, real groups generally do not have access to the material resources of their members beyond those resources invested in or gained from the group itself. For example, employers dissatisfied with their employees' performance can fire them or in some situations deduct fines from an employee's wages, but they cannot order employees to hand over their savings or their children's game console. This restriction is particularly relevant when we consider the broader social structure in which a group is embedded. In standard one-group experiments, a person's total payoff can easily be interpreted as resources gained from one group. When multiple groups are involved, this is no longer the case and a clear boundary needs to be in place between resources associated with the group and resources which are property of other groups and/or the individual. This restriction is implemented by our punishment mechanism. Second, our mechanism ensures that any points a participant does not contribute to either group are safe and not subject to punishment. This ensures that participants can avoid losses from one or both of the

groups they are a member of if they expect that this group will not be satisfied with any contribution they can make. This is particularly important in the two-group treatment with insufficient endowments. Imagine, for example, that a participant is a member of two groups which both punish any contribution lower than 30 points. Given the endowment of 40 points, there is no way for this participant to satisfy both groups, and this participant will likely receive punishment regardless of how they decide to distribute their endowment. If both groups could then distribute severe punishment not limited to the participant's gains from the group itself, it might be impossible for the participant to avoid overall negative returns in each round (and thus from the experiment as a whole) no matter their decisions.

We implemented the collective decision rule because the punishment distributed by this mechanism is quite severe, immediately reducing a punished participant's income from the group to zero, and because no additional punishment can be given to this participant once the exclusion has taken place.

Because this punishment mechanism is unusual compared to the existing literature on punishment in social dilemmas, we first tested the effectiveness of this punishment mechanism in an otherwise standard situation with only one group playing a PGG. The main purpose of the one-group treatment is therefore to demonstrate the effectiveness of the punishment mechanism. If the punishment mechanism is effective in the one-group treatment and we observe low levels of contributions in the two-group treatments, then we will know that this is due to the two-group structure and not due to the selected punishment system.

Procedure

The experiments were conducted at the Sociological Laboratory of the University of Groningen (http://www.soclab.nl). The Sociological Laboratory has a subject pool consisting of students at the University of Groningen. These students come from a variety of disciplines including sociology, economics, law, biology, physics, etc. Within the subject pool, psychology students and sociology students are overrepresented, compared to the population of students at the University. The rules of the Sociological Laboratory guarantee subjects that they will not be deceived in the experiment, and that they will be paid for their efforts.

Experiments took place in computer rooms prepared in such a way that, once they were seated, participants could not see the screens on which the other participants were playing. The experiments were programmed using the oTree framework for social science experiments (Chen, Schonger, & Wickens, 2016). The

experiment started with an introduction by the experimenter, explaining the rules of conduct within the lab and asking the participants to start reading the instructions. Instructions were provided on participants' computer screens, and during every step of the experiment, the relevant section of the instructions was available to participants on their screen. During the experiment, subjects were always allowed to take notes. This ensured that if participants wanted to remember information across periods (e.g. the past behavior of their peers) they were not required to memorize this information.

During the experiment, participants earned points depending on their decisions and those of their group members. At the end of the experiment, these points were converted to Euros at a fixed rate, such that on average participants earned around 8 euros from the experiment.

Data

Data were collected from a total of 112 participants across 7 sessions (32 participants across two pilot sessions, 80 participants across 5 regular sessions). Our original plan called for 12 sessions: two pilot sessions plus ten regular sessions. Of the 10 planned regular sessions, 2 were planned for the one-group treatment and 4 for each two-group treatment. This number of sessions proved infeasible due to constraints on the availability of participants. In the end, including pilot sessions, data were collected on three sessions of the one-group treatment and two sessions of each two-group treatment. The treatment to be played in each session was determined randomly before the first session.

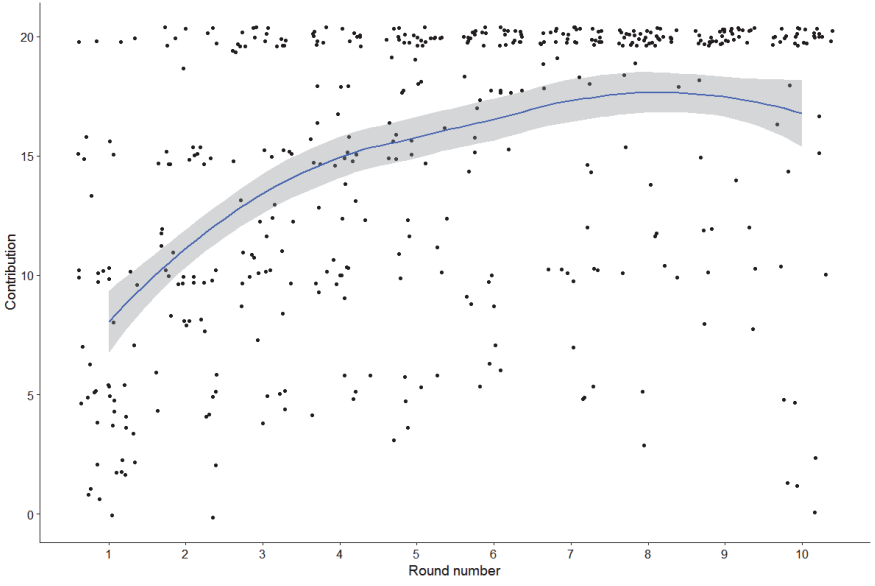
In our analyses, we will analyze the number of points *not contributed* by each participant in a particular round of the experiment. We will also analyze whether participants were punished in a particular round of the experiment. We have information on 1120 cases (112 participants during 10 rounds of the experiment) for each of these variables. Each case is cross-nested in an individual (the participant who took the decision) and two groups (the A- and B-groups this participant was part of at the time of the decision). This multilevel structure will be taken into account in all statistical analyses performed in the following section. We estimate multilevel models using the R package lme4 (Bates et al., 2015). In each analysis, we include uncorrelated random intercepts for participants, A-groups and B-groups, in addition to the level-1 error.

RESULTS

The effectiveness of the punishment mechanism

We tested the effectiveness of our consensus-based exclusion punishment mechanism in a standard Public Good Game. If the mechanism is effective we expect to see increasing contributions over time and punishments which are mainly directed at low contributors (Balliet et al., 2011; Fehr & Gächter, 2002; Oliver, 1980; Van Lange, Rockenbach, et al., 2014). We test the effectiveness in the baseline one-group treatment, which other than the punishment mechanism is identical to the one-group PGGs used in previous research. In two pilot sessions, the one-group treatment was played with partner matching. That is, participants played in the same group composition in all ten rounds of the experiment. In our main experiment, we conducted one additional session of the one-group treatment, with the stranger matching scheme that we also use for the two-group treatments. We inspected the partner and stranger treatments separately and found very similar results. Here, we present descriptive results which combine data from all three sessions, both partner and stranger matching.

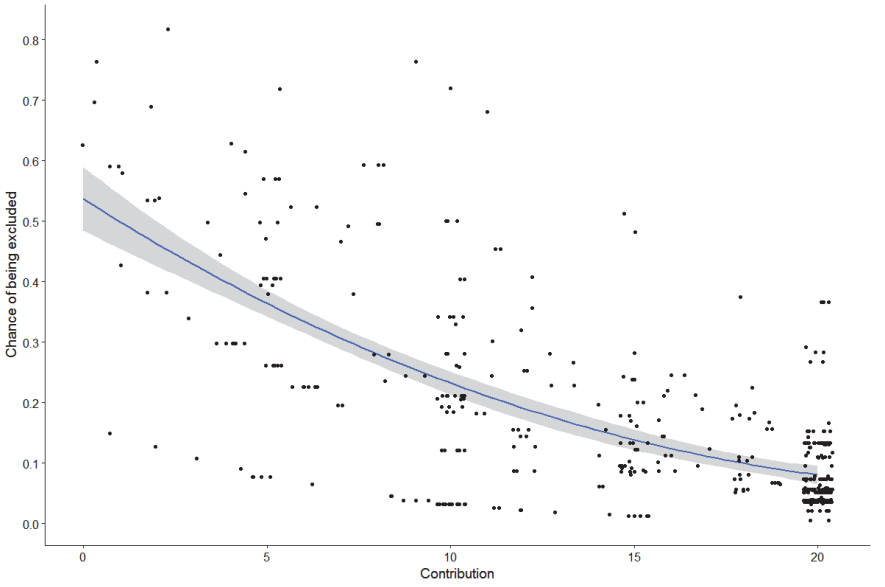
Figure 5.2. Contributions per round (jittered, with LOESS smoothing)



Note. The shaded area represents 95% confidence intervals

In total, we have information on 480 contribution decisions across the three sessions. In one of the sessions, a software error resulted in contributions higher than 20 being registered in some of the later rounds of the session. In total, this error affected 20 contribution decisions, which were excluded from our descriptive analyses. As can be seen in Figure 5.2, mean contributions start low (mean of 7.73 points in the first round) before rising up until the penultimate round (mean of 18.1 points in round 9) and falling a little in the last round of the experiment (mean of 16.4 points in round 10). The punishment mechanism appears effective at enforcing high contributions.

Figure 5.3. Predicted probabilities of received punishment (jittered, with LOESS smoothing)



Note. The shaded area represents 95% confidence intervals

To further support the conclusion that the punishment mechanism worked as expected, we look at the frequency with which punishment was distributed and the relationship between received punishment and contributions. Overall, there were 480 opportunities for exclusion (given that each participant could be excluded once per round of the experiment). Because we will examine the relationship between contributions and punishment, we once again excluded the 20 contribution decisions affected by a software error. Of the remaining 480 opportunities for punishment, exclusion occurred in 79 cases (17.17%). These punishments are mainly directed at individuals with contributions below the group mean (58 out of 79, 73.4%). To

illustrate the relationship between contributions and the probability of being punished, we estimated a multilevel logistic regression model[5] with uncorrelated random intercepts for participants and groups. The predicted probabilities of expulsion for each case, based on the coefficients of this model, are visualized in Figure 5.3. Lower contributions were more likely to trigger punishment.

Punishment of low contributions in two-group Public Good Games

Hypothesis 1 predicts that participants with low contributions are less likely to be punished when participants cannot contribute fully to all groups in which they participate. To test this hypothesis we investigate the frequency with which punishment happens and the relationship between received punishment and contributions in the two two-group treatments. Overall, we have 1280 opportunities for punishment across the two two-group treatments, whereby we treat each participant's exclusion (or lack thereof) from the A-group and the B-group as two separate cases. We find that punishment is more frequent when participants cannot contribute fully to all groups in which they participate (17.7% of all cases received punishment, versus 9.4% when participants can contribute fully to both groups). Mean contribution levels are higher in the treatment with insufficient endowments (18.8 points versus 17.5 points) but lower relative to the maximum contribution possible in each group (18.8 out of a maximum of 40 versus 17.5 out of a maximum of 20).

To formally test Hypothesis 1, we estimate multilevel logistic regression models with uncorrelated random intercepts for participants and groups. The dependent variable in these models is whether or not a participant was excluded from a particular group in a particular round of the experiment (0 = no, 1 = yes). The main independent variables are the number of points contributed by the participant (0 – 40), the treatment (0 = sufficient endowments, 1 = insufficient endowments) and the round of the experiment. In the second model, we also include an interaction term between the contribution and the treatment, to investigate whether similar contributions are less likely to receive punishment under the insufficient endowments condition.

Table 5.2 shows the results from these models. The predicted probabilities of expulsion for each case, based on the coefficients of this model, are visualized in Figure 5.4. We find that participants who contribute more to a group are significantly less likely to be excluded from the gains of that group (Model 1, b = -0.281, $p < 0.001$).

[5] More information on the model specification and parameter estimates of this model are available in an online appendix stored on the Open Science Foundation framework (https://osf.io/yhs98/?view_only=51a4a0141b154dc8b2c0613ee995ee14).

We also find that exclusion is *more likely* in the treatment where participants cannot contribute fully to all groups in which they participate (Model 1, $b = 1.012$, $p = 0.002$).

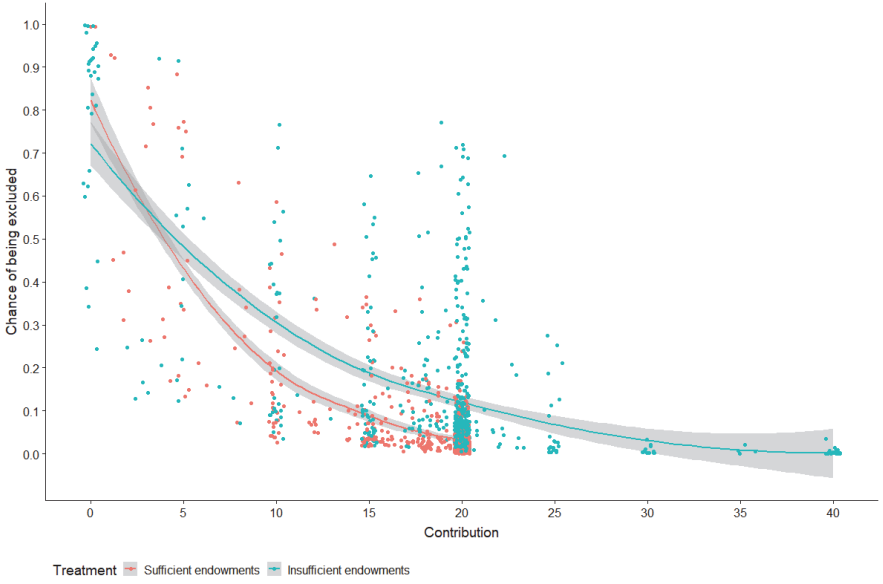
At first glance, this result contradicts Hypothesis 1, which predicted that exclusion would be *less likely* in this treatment. However, there is some nuance to these results. Model 2, in which we include the interaction between the participant's contribution and the treatment, shows that among very low contributions exclusion is, in fact, less likely in the treatment with insufficient endowments. Among higher contributions, however, exclusion is more likely in the treatment with insufficient endowments than in the treatment with sufficient endowments. This is evidenced by the interaction between contribution and treatment (Model 2, $b = 0.173$, $p < 0.001$). Because the scale of possible contributions is different in the treatment with insufficient endowments (0 – 40 points) than in the condition with sufficient endowments (0 – 20 points), we performed several additional analyses to investigate whether the nonexistence of contributions over 20 points in one of our treatments influenced these results.

Table 5.2. Estimates of multilevel logistic regression models for exclusion

	Model 1	*Model 2*
Treatment variables		
Contribution	-0.281 (0.034)***	-0.410 (0.057)***
Insufficient endowments	1.012 (0.320)**	-1.753 (0.850)*
Contribution x Insufficient endowments		0.173 (0.050)***
Control variables		
Round	0.276 (0.056)***	0.337 (0.064)***
Constant	0.084 (0.435)	1.745 (0.675)
Random effects (SD)		
Group	1.429	1.463
Participant	0.596	0.621
N	1280	1280
Deviance	803.7	790.1***

*Note. Two-sided p-values * < 0.05 ** < 0.01 *** < 0.001*

Figure 5.4. Predicted probabilities of exclusion (jittered, with LOESS smoothing)



Note. The shaded area represents 95% confidence intervals

When analyzing the two treatments separately, we find a negative relationship between the number of points contributed to a group and the probability of exclusion from that group in both the treatment with sufficient endowments (b = -0.383, se = 0.080, $p < 0.001$) and the treatment with insufficient endowments (b = -0.249, se = 0.038, $p < 0.001$). Comparing predicted probabilities of exclusion, we find that among very low contributions exclusion is more likely in the condition with sufficient endowments, while among higher contributions (even around 20 points) exclusion is more likely in the condition with insufficient endowments. When analyzing only those cases from both treatments in which the participant contributed 20 points or less to the group, we find the same interaction between contribution and treatment which we observe in the models presented in Table 5.2 (b = 0.157, se = 0.065, p = 0.016).

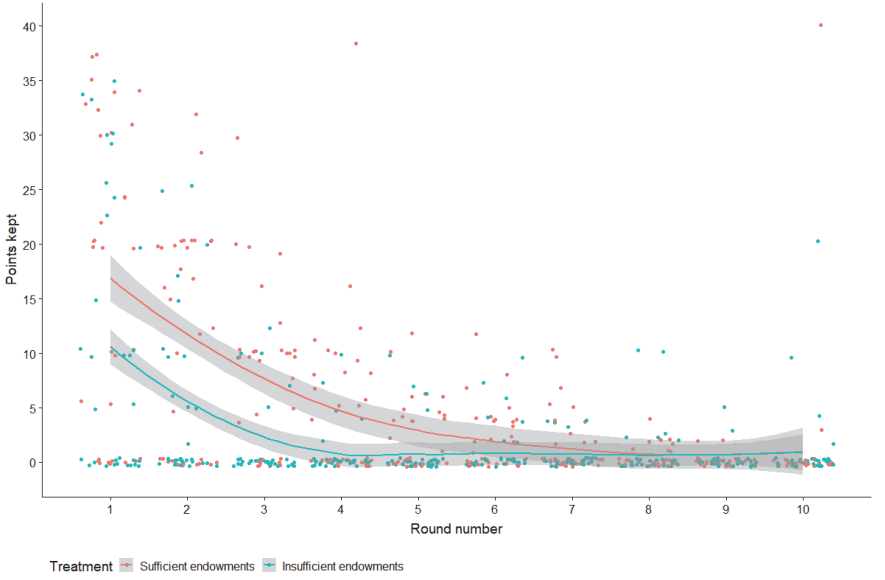
Overall, this gives us confidence that the results obtained were not influenced by the difference in the scale of the contribution variable between the two conditions.

Result 1: Very low contributions are less likely to result in exclusion when players cannot contribute fully to both groups. However, higher contributions are more likely to result in punishment when players cannot contribute fully to both groups.

Contributions in two-group Public Good Games

Hypothesis 2 predicts that contributions will be lower when participants cannot contribute fully to all groups in which they participate. Our results are not consistent with this hypothesis, and in fact, show the opposite. On average, participants keep fewer of their points in the two-group treatment with insufficient endowments (mean points kept = 2.32, *SD* = 6.10) than in the two-group treatment with sufficient endowments (mean points kept = 4.93, *SD* = 8.65). As can be seen in Figure 5.5, the difference is greatest in the first round of the experiment and disappears in the later rounds. In both treatments, participants keep very few points to themselves in the later rounds of the experiment.

Figure 5.5. Points kept by round of the experiment (jittered, with LOESS smoothing)



Note. The shaded area represents 95% confidence intervals

The distribution of the dependent variable, the number of points each participant kept for themselves in a particular round of the experiment, closely resembles a Poisson distribution. Estimating a multilevel linear regression model with normal errors results in severe violations of the assumptions of normality and homoscedasticity of residuals. To formally test Hypothesis 2, we therefore estimate multilevel Poisson models with uncorrelated random intercepts for participants, A-groups, and B-groups. The

dependent variable in these models is the number of points kept by a participant in a particular round of the experiment (0 – 40). The main independent variables are the treatment (0 = sufficient endowments, 1 = insufficient endowments) and the round of the experiment. In the second model, we also include an interaction term between the treatment and the round, to investigate whether the trend over time differed significantly between the two treatments.

Table 5.3 shows the results of these models. We find that participants keep significantly fewer points to themselves in the treatment with insufficient endowments than in the treatment with sufficient endowments (b = - 1.90, p = 0.003 two-sided). We also find that participants keep significantly fewer points to themselves in later rounds of the experiment (b = -0.55, p < 0.001 two-sided).

Result 2: Participants keep significantly fewer points for themselves when punishers cannot judge accurately whether non-cooperative behavior reflect non-cooperative intentions

Table 5.3. Estimates of multilevel Poisson models for points kept

	Model 1	*Model 2*
Treatment variables		
Insufficient endowments	-1.898 (0.636)**	-2.626 (0.853)**
Insufficient endowments x Round		0.148 (0.115)
Control variables		
Round	-0.550 (0.060)***	-0.619 (0.082)***
Constant	2.479 (0.523)	3.096 (0.782)
Random effects (SD)		
A-group	1.298	1.284
B-group	1.273	1.269
Participant	2.072	2.059
N	640	640
Deviance	2535.2	2533.6

*Note. Two-sided p-values * < 0.05 ** < 0.01 *** < 0.001*

CONCLUSIONS

As a first step towards exploring the understudied influence of overlap in group membership on behavior in social dilemmas, we applied findings on the monitoring of free-riders to a multiple-group situation with overlap in group membership. When multiple groups share members, and behavior of fellow group members in other groups cannot be monitored, judging the selfishness of others' behavior is difficult.

Based on the knowledge that punishment is generally directed at free-riders, who are judged to have non-cooperative intentions, we expected that participants would be reluctant to punish fellow group members when judging their intentions was difficult. This expectation was supported by findings from previous studies in single groups. Imperfect monitoring makes it difficult to accurately identify non-cooperative group members, thus hindering the effective distribution of punishment (Patel et al., 2010; Van Miltenburg et al., 2017).

As expected, we found that very low contributions were less likely to be punished when punishers could not judge whether low contributions reflected non-cooperative intentions. Low contributions, in particular, are difficult to judge in such a situation, as they can reflect both full free riding (contribution next to nothing in both groups) and full cooperation in one of the groups (which the other group would observe as a very low contribution).

At the same time, we found that punishment of higher contributions is *more* common in the treatment where punishers could not accurately judge intentions from behavior. This is likely to be a result of the fact that contributions were capped at 20 points per group in the one treatment, while they could go all the way up to 40 points in the other treatment. Thus, in the treatment with sufficient endowments, a contribution of 20 points was the maximum contribution while in the treatment with insufficient endowments a contribution of 20 points was only half of the maximum. It appears that the possibility of higher contributions did raise the bar for acceptable contributions somewhat. Some group members may have wanted the other in their group to invest more than 20 points, to the possible detriment of contributions in other groups.

When we look at the actual contributions which were made we find that even in the treatment with insufficient endowments the vast majority of contributions is close to 20 points. Participants did not keep more points to themselves when they could disguise this action as contributions to another group, perhaps mainly because even in the treatment with insufficient endowments low contributions were sufficiently likely to result in exclusion. The groups appear to have coordinated on sharing the points equally.

Overall the results, although they initially appear mixed, actually paint a fairly consistent picture. To some extent, punishers did show more restraint in punishing low contributions when these low contributions could be explained as high contributions in another group. However, this did not lead to significantly more free riding. An explanation for this result may be that an assumption underlying this

expectation, namely that very high contributions in another group would be considered cooperative behavior, does not entirely hold. As our results show, extremely high contributions are very rare and an equal division of resources across the two groups is far more common. Low contribution, then, represent a deviation from the (descriptive) norm. It may well be that most participants considered an equal division across the groups to be the suitable cooperative behavior.

DISCUSSION

Suggestions for further research

Our intention was also to introduce the topic of overlap in group membership, and more broadly the social structure in which groups are embedded, as an important factor in social dilemmas. There are real opportunities to translate research on social dilemmas to practical applications, and one of the major barriers to this translation may be the neglect of social structure social dilemma research. Social dilemmas rarely take place in isolation. Other groups create constraints on individual behavior, are subject to externalities generated by solving social dilemmas within a group, and are the source of competing pressures. The willingness to cooperate which is present in many individuals has been established (Balliet et al., 2009; Pletzer et al., 2018). We know more and more about how we can encourage cooperation (Chaudhuri, 2011; Fehr & Gächter, 2002), and what the limits of these strategies may be (Guala, 2012). Now, the question facing us is not 'do people cooperate?' but 'where do people cooperate?'. Given limited resources and a delicate balance between the interests of multiple groups, to which groups do people choose to contribute?

In the present study, we have given one example of how this new area of research may be approached. By drawing on existing literature about social dilemmas, and applying it to a situation with multiple intersecting groups, we learn about the generalizability of established findings. For instance, while implementing our peer punishment institution with collective decision rule in one-group public good games leads to convergence on the maximum possible contribution, the same rule leads to a fair division between groups in a two-group public good game with competing interests. Others may apply knowledge about the importance of social structure from fields such as social identity theory (Tajfel, 1982; Tajfel & Turner, 1979) or social network research, to social dilemmas. Yet others may use the social structure in which a social dilemma is embedded as predictors of relevant factors in social dilemma behavior, such as the different norms with which individuals enter a social dilemma

situation (Rauhut & Winter, 2017; Winter et al., 2012). In our design, the fact that all participants were in the same structural situation (as a member of two groups which share no other members) may have influenced the observed result, in which endowments are distributed equally across both groups.

Finally, there are open questions which relate specifically to situations in which multiple groups intersect. For instance, when groups compete for an individual's resources, what determines which group wins out? Are individuals able to sustain membership in groups with competing interests? How do people reconcile competing demands? These questions are highly relevant to many situations in which social dilemma research could be applied. Think, for instance, of teams working in an organization. The team is a subgroup of an organization, and team members may have tasks to perform in several teams at once. Given the ubiquity of multiple (overlapping) group membership in real-life social dilemma-like situations, investigating these question is an important task for future studies.