Cooperation between newcomers and incumbents: The role of normative disagreements

Abstract

Cooperation in groups often requires individual members to make costly contributions that benefit the group as a whole. Prior research suggests that shared norms can help to support ingroup cooperation by prescribing common standards of how much to contribute. These common standards may be disrupted when groups undergo membership change, i.e., when members from outgroups enter the ingroup. When newcomers and incumbents have different notions about how much to contribute, a normative disagreement ensues that could undermine cooperation and the extent to which individuals identify with the group. In a laboratory experiment, we manipulate whether newcomers and incumbents disagree about how much to contribute in a public goods game with peer punishment. We examine whether normative disagreement between newcomers and incumbents affects newcomer-incumbent relations in terms of group identification, the emergence of a social norm, and costly punishment. The main goal is to test whether normative disagreement and the resulting newcomer-incumbent relations harm cooperation in terms of contributions to the common good. We find that normative disagreement between newcomers and incumbents negatively affects the emergence of a shared social norm, leads to more costly punishment just after membership change, and lowers feelings of group identification. Contrary to expectations, normative disagreement does not affect cooperation negatively. Instead, norm enforcement via punishment causes participants to adjust their behavior to each other's standards. This norm enforcement is especially directed at newcomers, leading them to conform to the incumbents' standards.

Introduction

Group cooperation often requires individual members to make costly contributions that benefit the group as a whole. For example, countries provide public and social security based on citizens' tax payments, neighborhoods maintain clean and safe parks if residents abstain from littering and keep watch, and work organizations survive and grow as a result of collaboration between workers (Dur & Sol, 2010; Sanders, 2009; van Gerwen et al., 2018; Wageman, 1995). The composition of such groups changes frequently due to the arrival of new members and departure of old members. For example, work organizations hire new workers and let go of existing workers who retire or move to other organizations; countries, cities, and neighborhoods change in composition due to migration, and volunteer organizations and cooperatives attract new members and see other members leave. Sustainable group cooperation thus requires that contributions to the common good continue, regardless of the turnover in group members. However, this cooperation is often theorized to be impeded by newcomers and incumbents having different notions about how much should be contributed to the common good (Collier, 2013; Habyarimana et al., 2009; Ostrom, 2000), a situation which we will refer to as normative disagreement. In this study, we examine experimentally whether normative disagreement between incumbents and newcomers harms cooperation in terms of contributions to the common good. We also explore how perceptions of the contribution norm and feelings of ingroup identification shift due to the arrival of newcomers in the group.

There is a vast body of research that studies cooperation experimentally in the lab using social dilemma games (Chaudhuri, 2011). Most of this research is directed either at intragroup cooperation or intergroup cooperation. A main finding is that intragroup cooperation can be sustained via the development and enforcement of contribution norms (Fehr & Schurtenberger, 2018a). Although these norms promote intragroup cooperation, they can impede intergroup cooperation (De Dreu et al., 2020). Norms are typically group-specific and favor the ingroup over the outgroup (Bernhard et al., 2006; Titlestad et al., 2019). This can lead to conflict when different groups have to cooperate together but each group wants to stick to their own norm (Jetten et al., 1996). Indeed, there is ample research showing that intergroup relations are often characterised by conflict rather than cooperation (Balliet et al., 2014; Böhm et al., 2020). While research using social dilemma games has advanced our knowledge about intragroup and intergroup cooperation, we know much less about cooperation in intermediate cases where group affiliations are in flux. In particular, we know little about cooperation in groups where newcomers enter and have to cooperate with incumbents to contribute to the common good.

Newcomers are often seen as outsiders by incumbents (Rink et al., 2013) and newcomers do not readily identify with incumbent-groups (Moreland, 1985). Newcomer-incumbent relations are thus a special case between intragroup and intergroup relations that have received little attention in social dilemma research.

The few studies that have examined the influence of newcomers on cooperation in social dilemma games mainly looked at overall effects, i.e., whether newcomer entry has a positive or negative effect on contributions to the common good. The results are mixed. Some studies suggest a positive effect of newcomers on contributions to the common good (e.g., Duffy & Lafky, 2016; Sonnemans et al., 1999), whereas other studies report a negative effect (e.g., Grund et al., 2015; Ranehill et al., 2014). As far as we know, this prior literature has not examined potential underlying mechanisms producing these effects, including the common theoretical prediction that normative differences between incumbents and newcomers harm cooperation. The failure to cooperate as a result of people holding incompatible normative views has been referred to as a normative conflict (Winter et al., 2012). We thus examine to what extent normative disagreement between newcomers and incumbents leads to conflict in terms of cooperation failure.

Stability in group composition is commonly theorized to promote contributions to the common good via shared social norms (Fehr & Schurtenberger, 2018a). Repeated interaction with the same group members facilitates the reaching of a social norm, i.e., a common understanding of what is an appropriate contribution level (Duffy & Ochs, 2009). This in turn helps group members to know what to expect from others and hold one another accountable for uncooperative behavior. Because norms are often group-specific, different groups develop different norms on how much to contribute (Bernhard et al., 2006; Gangadharan et al., 2017; Henrich et al., 2001). When members migrate between groups that hold different and incompatible norms, normative disagreement ensues. When members instead migrate between groups that hold similar norms, the arrival of newcomers does not introduce normative disagreement. If a shared understanding of appropriate behavior is indeed an important element of cooperation, the impact of membership change on cooperation could depend on the level of normative disagreements between incumbents and newcomers.

We make use of the public goods game (PGG) to examine to what extent normative disagreement between incumbents and newcomers causes cooperation to unravel. In our

laboratory experiment, participants are sorted into groups and have to decide how much to contribute to a group project. Contributions to the group project are individually costly but give a return to all members in the group. Collective payoffs are maximized when all members contribute fully to the group project whereas individual payoffs are maximized by free-riding on the contributions of others. This elementary decision situation constitutes a social dilemma and is arguably reflective of many real-life instances of group cooperation (De Cremer & van Knippenberg, 2003; Dur & Sol, 2010; Sanders, 2009; van Gerwen et al., 2018). We study a PGG where the return rates of the group project differ between group members. PGGs with heterogeneity in return rates induce more interpersonal variation in normative views about how much to contribute to the group project (Reuben & Riedl, 2013), which allows us to manipulate normative disagreement as we will show.

We manipulate the level of normative disagreement between incumbents and newcomers in two conditions of the PGG as illustrated in Figure 1. In both conditions, we first let participants in groups of three play 10 rounds of the PGG together. Such repeated interactions with the same set of participants have been shown to foster ingroup favoritism and the development of group-specific norms (Dorrough et al., 2015; Titlestad et al., 2019). To indicate the group's identity, each group is randomly assigned a color at the start (orange or blue). After the first 10 rounds, we replace one member per group for a member from another group with another color (the newcomer) before letting them play another 10 PGG rounds. Before letting participants play the PGG, we measure each participant's view on how much each member should contribute (their normative view). The conditions differ in how we sort and resort participants into groups based on their normative views. In one condition, we sort participants for the first 10 rounds into groups that agree on the appropriate contribution towards the group project. Before starting the second 10 rounds, we replace an existing member by a new member who disagrees with the incumbents of the group about the appropriate contribution. Groups in this condition thus move from agreement to disagreement due to membership change, which is why we call this condition agreement-disagreement. In the other condition, we sort participants for the first 10 rounds into groups where one member disagrees with two other members. Before starting the second 10 rounds, we replace this disagreeing member with a new member who does agree with the other two members. Groups in this condition thus move from disagreement to agreement due to membership change, which is why we call this condition disagreement-agreement. As we will show, normative disagreement indeed increases due to membership change in condition agreement-disagreement, while it decreases in condition disagreement-agreement. We

4

hypothesize that contribution levels decrease more after membership replacement if the replacement results in more disagreement.

Before we test the impact of normative disagreement on contribution levels, we examine how normative disagreement affects the development of newcomer-incumbent relations. The influence that normative disagreement has on cooperation may depend on the type of newcomer-incumbent relations that develop under normative disagreement. We examine three dimensions of the newcomers-incumbent relations. We examine whether normative disagreement between newcomers and incumbents affects (1) group identification, (2) costly peer-punishment, and (3) the emergence of a social norm (i.e., convergent normative expectations on how much should be contributed to the common good).

The difference in contribution levels between conditions *before* membership change has been reported in another paper (Otten et al., 2020). We found that normative disagreements did not influence cooperation in newly formed groups (i.e., groups with no prior interaction). Instead, group members were mostly tolerant of other members contributing according to different normative views. In the current paper, we look at the contribution levels *after* membership change, to examine whether normative disagreements between newcomers and incumbents impede cooperation. As we will show, normative disagreement between newcomers and incumbents has substantially different consequences than normative disagreement in groups where all members are new. Our results, as we will report below, indicate that normative disagreements between newcomers and incumbents harm group identification (especially among newcomers), lead to more costly punishment just after membership change (especially directed at newcomers), and impede social norm emergence, all of which does not occur in newly formed groups. Nevertheless, newcomers and incumbents cooperate for the common good. Participants that contribute low amounts are punished and react by adjusting their contribution levels upwards. However, the contribution levels that groups eventually reach mostly reflect incumbent norms. This implies that there is an unequal distribution of normative influence with newcomers having less influence than incumbents.

We make three contributions to the literature. First, we study cooperation between newcomers and incumbents, a type of cooperation positioned between intragroup and intergroup that has rarely been researched in experiments with social dilemma games. Second, we examine how normative disagreements impact the development of newcomer-incumbent relations in terms

of group identification, costly peer-punishment, and social norm emergence. Finally, we examine whether normative disagreement between newcomers and incumbents causes conflict in terms of cooperation failure.

Figure 1. Experimental design.
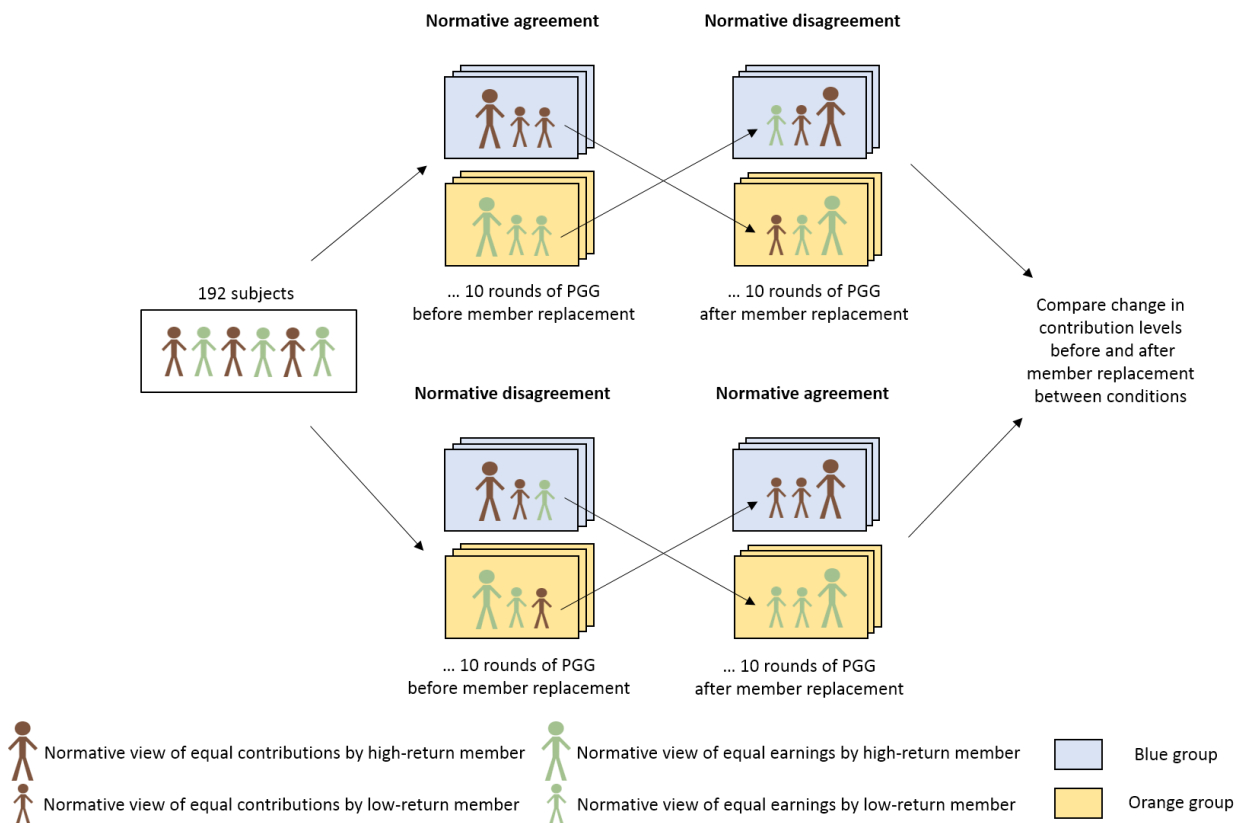
Literature review

Membership changes in public good games

One of the earliest experiments on membership changes in the PGG is conducted by Sonnemans et al. (1999). The study examines 4-player groups in which one randomly selected member is replaced by another after a prespecified number of rounds. Group composition and the schedule of group changes are common knowledge from the start. The authors find that participants decrease their contribution just before they leave the group, but in the new group substantially increase their contribution. As a result, contribution levels increase considerably just after each membership change. Similarly, Duffy and Lafky (2016) find that periodically replacing old members by new members helps 4-player groups to sustain contribution to the public good over a longer period of time compared to groups of stable composition.

However, other studies report negative effects of membership changes on contribution levels. Grund et al. (2015) examine contribution levels in blended 4-player groups, where some group members stay together for all rounds (partners) and other members switch groups every round (strangers). There are four conditions that differ in whether the groups consist of (1) all partners, (2) three partners and one stranger, (3) two partners and two strangers, or (4) all strangers. The authors find that the contribution level is lower in groups with more strangers (i.e., more membership changes). Ranehill et al. (2014) study how the rate of newcomer entry affects contribution levels in growing groups. They find that a higher rate of newcomer entry (i.e., more newcomers at once vs each newcomer entering in separate rounds) negatively affects contribution levels. Finally, McCarter and Sheremeta (2013) find that, in a minimum-effort game, newcomers have a negative effect on the effort devoted to group cooperation. There is a related literature on PGGs where members themselves can form groups and thus also have the option to enter and leave groups. These studies are designed to examine group formation processes rather than the effects of membership changes, as causal inferences about membership change effects are difficult to draw when members can select themselves into groups (i.e., selection effects). For our purposes, these studies are thus of less interest, but a recent literature overview can be found in Guido et al. (2019).

We have seen that studies that examined the effects of membership changes generally looked at overall effects, i.e., whether newcomers affect contributions to the public good, and produced mixed results. Potential mechanisms that drive these newcomer effects and could explain these conflicting findings have received much less attention. As mentioned, our study focuses on the role of normative disagreements between incumbents and newcomers. In the next section, we turn to the literature that suggests how newcomer-incumbent relations can be regarded as a special case positioned between intragroup and intergroup relations and how this depends on normative disagreement between newcomers and incumbents.

Intergroup differentiation in newcomers-incumbents relations
When examining the relevance of intergroup research for newcomer-incumbent relations, a first question that arises is whether incumbents categorize newcomers as ingroup or outgroup and vice versa. According to the group socialization model (Levine & Moreland, 1994), newcomers are initially not seen as full ingroup members by the incumbents. Instead, newcomers occupy a position between non-members and full members. Only after newcomers have experienced a

socialization process, during which incumbents attempt to get the newcomers to act in line with the group's goals and norms, do the newcomers become full members. Thus, only over time newcomers are said to make a transition from 'outsiders to being insiders' (Bauer et al., 2007). In support of this group socialization model, most empirical evidence suggests that newcomers are initially regarded as outsiders by incumbents (Rink et al., 2013). Similarly, newcomers do often not immediately identify with the incumbent-group (Moreland, 1985). Newcomers' identification with the incumbent group may be especially low when the membership change is not initiated by the newcomers themselves but rather by an external decision-maker, such as the experimenter as in our study (Arrow & McGrath, 1993).

Self-categorization theory explains when and why people consider themselves as members of a particular group (Turner et al., 1987), and can therefore be used to predict when people are likely to identify with a group despite the presence of newcomers. According to this theory, one of the aspects that makes a set of people be seen as a self-relevant group is normative fit. When assessing normative fit, people compare a potential member's attributes and behaviors with one's expectations about dimensions that should distinguish between members of different groups in a particular situation. Thus, whether incumbents and newcomers contribute according to similar or different normative views will matter for whether they see themselves as one group or two groups. If newcomers act according to different normative views than incumbents, normative fit will be lower. This, in turn, will lead to a lower chance that newcomers are categorized as ingroup by incumbents.

Is categorization into different (sub)groups between incumbents and newcomers able to lead to conflict in terms of cooperation failure? According to social identity theory, ingroup favoring biases that impede collective cooperation may emerge once people are categorized into different groups (Tajfel & Turner, 1986). This has been demonstrated in several situations and expressions of bias, even where categorization into ingroups and outgroups is based on arbitrary criteria (see for a meta-analysis of economic experiments: Lane, 2016). Nevertheless, cooperation failure is especially likely when newcomers threaten the group's norms and goals (Böhm et al., 2020; Thravalou et al., 2020) Newcomers that contribute in line with a normative view that conflicts with the normative view of incumbents bring a larger threat to both the group's norm and the group's goal in terms of realizing the common good. Consequently, normative disagreements between incumbents and newcomers can not only matter for how newcomers are categorized, but also for the likelihood of cooperation failure.

In short, newcomers and incumbents do not readily identify with each other, and instead need a socialization process before they feel they belong to the same group. When newcomers are not regarded as part of the ingroup by the incumbents and vice versa, cooperation failure becomes more likely, especially when newcomers and incumbents are in normative disagreement. We next review how the effect of normative disagreement on cooperation in public goods games has been studied so far.

Normative disagreement in PGGs

In the typical PGG, all members can contribute similar amounts and obtain similar benefits from the public good. In these circumstances, a norm of equal contributions by all members is commonly supported and enforced among participants in these experiments (Carpenter & Matthews, 2009; Fehr & Fischbacher, 2004; Fehr & Schurtenberger, 2018a). When groups are heterogeneous in terms of their members' contribution capacities or the returns from the public good, there is more interpersonal variation as to which contribution norms are supported. With heterogeneity in the returns from the public good, for example, there are at least two norms that are commonly supported by different participants (Reuben & Riedl, 2013). The first norm is referred to as equal-contributions and prescribes that all actors contribute equally to the public good. The members with a higher return from the public good then end up with higher earnings. The second norm is referred to as equal-earnings and prescribes that high-return members contribute more than low-return members, such that the earnings are equalized. A balance between both norms is also possible, i.e., norms that prescribe that high-return members should contribute more than others, but not to such an extent that earnings are equalized (Otten et al., 2020).

Whether the variation in normative views about PGGs with heterogeneous returns is also reflected in terms of actual contribution behavior depends on whether norms can be enforced. Norm enforcement in PGGs most often takes the form of peer punishment, where participants punish norm breakers by deducting points from their payoffs (an alternative norm enforcement mechanism is rewarding norm followers by assigning bonus payoffs). Without means to enforce norms, most groups show a trend to free-riding regardless of any between-group differences in normative views (Reuben & Riedl, 2013). However, if participants can enforce norms, clear differences between groups emerge. Some groups contribute in line with the equal-

contributions norm, others with the equal-earnings norm, and yet others with a balance between the two (Gangadharan et al., 2017; Otten et al., 2020).

The importance of norm enforcement for sustaining prosocial contribution norms has been demonstrated in many studies (Chaudhuri, 2011; Fehr & Gachter, 2000). However, sanctioning does not always have a cooperation-enhancing effect. Sanctioning is sometimes also directed at people that made large contributions to the public good, a phenomenon known as antisocial punishment (Herrmann et al., 2008). If those who make a low contribution feel that their contribution is sufficient and nevertheless are sanctioned, they may refuse to increase their contribution and retaliate against high contributors, because these high contributors are the one's likely to punish the low contributors. Thus, for sanctions to have a cooperation-enhancing effect, it seems important that group members agree on what constitutes the appropriate contribution to the public good.

Because there is more variation in normative views in public good games with heterogeneous returns, there is also more potential for within-group normative disagreements. The prevailing conjecture is that normative disagreements harm cooperation in terms of contributions to the public good (Kingsley, 2016; Nikiforakis et al., 2012; Rauhut & Winter, 2017; Winter et al., 2012). In these studies on normative disagreement, participants do not have direct information on each other's normative views, but they do have information on each other's contributions. The effect of normative disagreement works via observing each other's contributions and judging whether these contributions match one's own normative view. Many people contribute only if their group members also contribute their share, i.e., they cooperate conditionally on others' cooperation (Chaudhuri, 2011; Thöni & Volk, 2018). If people differ in their normative views and contribute according to their own view, they will observe that their view is not adhered to by others. The expected consequence is that people who feel others are not contributing enough will reduce their own contribution, causing a downward trend in contribution levels.

The first study to manipulate normative disagreement experimentally found no evidence for a negative effect on contribution levels in groups where all members are new (Otten et al., 2020). However, such groups are substantially different from groups consisting of incumbents and newcomers with both having prior group affiliations and experiences. In groups with all new members, these members are initially unaware of what others deem appropriate contribution

behavior. In the course of interacting with each other, a common standard of behavior emerges and turns into a group-specific norm that members become accustomed and attached to (Diekmann & Przepiorka, 2016; Titlestad et al., 2019). Once in place, norms are not easily changed. Many experiments show that participants often keep conforming to a norm even if the incentive structure changes in favor of norm transgression or movements to a new norm (Andreoni et al., 2020; Duffy & Lafky, 2019; Guala & Mittone, 2010; Smerdon et al., 2019). Furthermore, participants take the norms they have learned in prior interactions with them when entering new social situations (Engl et al., 2017; Peysakhovich & Rand, 2013; Stagnaro et al., 2017), although not indefinitely (Arechar et al., 2018; Duffy & Ochs, 2009; Fréchette & Yuksel, 2017).

Incumbents do share a common history in which they have developed joint normative standards, and may thus be less open to conflicting normative views from newcomers compared to members lacking prior experience and group affiliation. Likewise, newcomers take the norms they have become accustomed to in their prior group with them when they enter another group. Even if newcomers wish to conform to the incumbents' norm, when the newcomers are accustomed to a different norm than the incumbents, they may have incorrect expectations about the incumbents' norm and consequently fail to conform to it. What is more, the aforementioned categorization processes that distinguish between newcomers and incumbents on the basis of normative views may also obstruct cooperation more in the presence of normative disagreement rather than agreement. Indeed, that newcomers are accustomed to different norms than incumbents is often argued to harm cooperation for the public good (Collier, 2013; Habyarimana et al., 2009; Ostrom, 2000). In sum, we hypothesize that membership change harms cooperation in terms of contributions to the public good more when newcomers and incumbents are in normative disagreement rather than agreement.

Methods

We conducted a computerized experiment in the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University with 192 participants sorted into 64 groups of 3 members each. We recruited participants amongst students at Utrecht University using the internet recruitment system ORSEE (Greiner, 2015) during October-November 2019. We conducted 8 sessions. Each session included 24 participants and lasted about 75 minutes. Payment depended on behavior in the game. On average, participants earned 15 euros (min = 5, max = 22). The average age of participants was 24 years, 127 (66%) were female, 62 male,

and 3 other. Almost all participants were attending courses at Utrecht University, 87 were of Dutch nationality, and 105 from various other countries. Participants were randomly placed in an individual cubicle and informed about the experiment through written instructions (provided in the supplementary material). The main part of the experiment consists of repeated rounds of a version of the public goods game with peer punishment (Fehr & Gachter, 2000). We first describe this game in its standard form and then outline the procedure that we used to adapt it for our purposes.

Game

Each game round has two stages. First, each individual i receives an endowment of 20 monetary units (MU) and decides how much to contribute to a public good, c_i, where $c_i \in \{0, 1, \ldots, 20\}$. The part of the endowment that i does not contribute is kept for oneself. The public good consists of the sum of the contributions made by the three members of the group. Each member receives a return m_i per point contributed to the public good, with $m_i < 1$. The sum of these returns is the multiplication factor of the public good M, with $N > M > 1$. Because $m_i < 1$, it is most profitable for the individual group member to contribute nothing in each round. However, since $M > 1$, it is most profitable for the group if every group member contributes their full endowment. These two aspects together constitute the social dilemma, i.e., the conflict between individual and collective interests, of public good provision. After all group members made their contribution decisions, the contributions and payoffs of each member are communicated to all group members.

Second, each group member is given the opportunity to assign punishment points $p_{ij} \in \{0, 1, \ldots, 10\}$ to each group member $j \neq i$. Each punishment point costs 1 point to the punisher, and reduces the payoff of the punished group member by 3 points. This provides participants with an informal instrument for norm enforcement (Fehr & Schurtenberger, 2018; Reuben & Riedl, 2013). The individual payoff (π_i) after one round of this two-stage game is calculated as follows:

$$\pi_i = 20 - c_i + m_i \sum_j c_j - \sum_{j \neq i} p_{ij} - 3 \sum_{j \neq i} p_{ji}$$

As is common in PGGs with peer punishment, we do not let participants see who punished whom. This curbs punishment driven by revenge motives instead of dissatisfaction with others' contributions and thereby helps to analyse punishment as an instrument for norm enforcement.

As described in the review section, there is more variation in normative views between participants when there is heterogeneity in the returns of the public good. As a consequence, there is a higher likelihood that different groups develop different contribution norms (Gangadharan et al., 2017), which increases the chance of normative disagreements when members are switched between groups. Per group of three members, we randomly assign two members a low return of $m_i = .50$ from the public good and one member a higher return of $m_i = .75$ (this makes the joint multiplication factor $M = 1.75$). Participants know which members have the low-returns and which the high-return. A prior study suggests that with this level of heterogeneity, people vary considerably in their normative views (Reuben & Riedl, 2013). For comparability to previous research, all other parameter values are set to follow the typical form of the PGG with peer punishment (Fehr & Gachter, 2000).

Procedure

We implement two conditions, that differ only in the method of sorting and resorting participants in groups based on their normative views on the spectrum of equal-contributions vs equal-earnings. In the first condition (agreement-disagreement), we initially sort participants from the same side of the spectrum together, whereas in the second condition (disagreement-agreement) we initially sort participants from different sides of the spectrum together (see Figure 1). After this sorting, each group receives a color (orange or blue) and the participants play an initial 10 rounds of the public goods game within their group. Participants are told that there will be a second part of the experiment after these 10 rounds and that they will receive information about this second part just before it starts.

After this first part of the experiment, we exchange one member per group for a member from another group with another color in both conditions, such that each group receives a newcomer in place of an old member. We inform the participants of this membership change and let the newly formed groups play a second set of 10 rounds. The resorting is done in such a way that the disagreement in normative views reverses between conditions; the groups in the condition with initial sorting on similar normative views become as dissimilar (in terms of initially elicited normative views) as the groups in the condition with initial sorting on dissimilar normative views, and vice versa. Thus, in one condition the newcomer increases the level of normative disagreement, and in the other condition the newcomer decreases it. To prevent experimenter demand effects, we did not inform participants about the method of group (re)sorting. We did

not use deception, i.e., we did not offer untruthful information to the participants and they were aware that they did not have information on how group sorting happened.

As is common in studies on normative disagreement in PGGs, we did not show participants each other's normative views. This helps to isolate the impact of normative differences from potential confounders such as expectations on what one's group members will contribute. Participants are thus initially not aware of how much they agree or disagree with their group members, and instead infer this indirectly from their contribution and punishment decisions. Participants are generally quite good at guessing other's normative views; 56% guess correctly on average before the experiment, and 71% guess correctly on average after the experiment (see the upcoming section on normative expectations for how participants guess others' normative views).

Because there are two low-return members and one high-return member per group, we switch low-return members between groups. This allows us to compare the behavior of the low-return newcomer with the low-return incumbent when doing individual-level analyses, i.e., to prevent confounding of newcomer-incumbent differences with return-rate differences. The exact method of sorting and resorting is outlined in Figure 2. The experiment is pre-registered at https://osf.io/gy8st.

Figure 2. Example for Method of Sorting and Resorting Participants
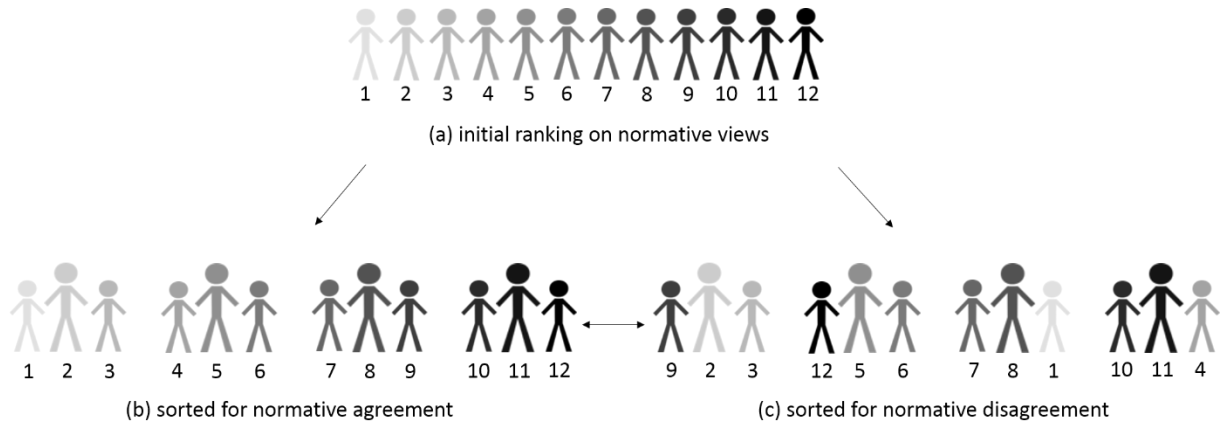
Note: At the beginning of the experiment (a), participants are ranked in terms of their normative views on the spectrum of equal-contributions vs equal-earnings (indicated with numbered grey shading). In the example presented here, there are 12 participants sorted into 4 groups. When sorting for normative agreement (b), we first form a group of the three highest-ranked participants (1-3), then of the remaining participants we again form a group of the three highest-ranked participants (4-6), and so on until all participants are grouped. When (re)sorting such that groups become dissimilar (c), we select the highest-ranked low-return participant from the first group in the first half of the groups (ordered in terms of support for equal-earnings over equal-contributions) and replace the participant with the lowest-ranked low-return participant from the first group in the second half of the groups, and repeat this procedure with the remaining groups. In this way, the extent of normative disagreement (in terms of rank-differences) is equal for all groups. In one condition participants start in groups sorted on similar normative views and then members are exchanged such that groups are sorted on dissimilar normative views (b → c), and in the other condition we reverse this order (c → b). In both conditions, there is one member per group that obtains a higher return from the public good than the other two members, as indicated by the size of the figures.

Normative views

Before playing the game and assigning the individual returns, we measured participants' normative views. To do so, we showed participants a hypothetical group of three members, two of which obtain a low-return ($m_i = 0.50$) and one of which obtains a high-return ($m_i = 0.75$), the exact same composition of returns as used in the actual contribution rounds of the experiment. We subsequently asked: "According to you, what is the appropriate amount that each member should contribute to the group account". Participants could then indicate a contribution for each of the three members between 0 and 20. Participants can try out different combinations of contributions, and see how it affects the earnings of each group member (see instructions and screenshots in the supplementary material). The normative view is measured again after the first 10 rounds of the game, and also one final time after the second 10 rounds of the game. Before this second and third measurement, we explicitly remind participants that they need not be consistent between the different measurement moments.

Participants' normative views are used to position them on a spectrum from equal-contributions to equal-earnings. Supporters of equal-contributions would answer that both types of players should contribute equally to the public good (ratio of contributions by high-return members to

low-return members = 1), whereas supporters of equal-earnings would answer that high-return types should contribute twice as much as low-return types (ratio of contributions by high-return members to low-return members = 2). Participants who support a balance between both rules would answer that high-return types should contribute more than low-return types, but not twice as much. As we will show, almost all of our participants fall within one of these three categories and are rather evenly distributed across the three categories.[1]

Normative expectations and social norms

After participants report their personal normative view, we tell them that their group members were also asked to indicate appropriate contributions for three members in the PGG. Each participant is then asked to guess the answers submitted by their group members, i.e., to report their normative expectations. To incentivize the guesses, participants are informed that one of their guesses will be picked randomly and yield an additional payment of 100 MU (~€1.40) if it matches the actual answer of at least one of the group members. Only at the end of the experiment, participants are informed of whether they were correct in the guess we randomly chose. This measure is inspired by earlier studies (Krupka & Weber, 2013; Reuben et al., 2015).

Many definitions of social norms share the view that social norms involve shared expectations between group members on what actions are considered appropriate (Fehr & Gachter, 2000; Ostrom, 2000). As a consequence, mutually consistent normative expectations in a group are commonly used as indicator of the existence of a social norm (Bicchieri, 2006; Bicchieri et al., 2014; Krupka & Weber, 2013). Therefore, to assess the existence and emergence of social norms in the experiment, we examine the overlap in normative expectations between members within groups. As mentioned, all participants were asked to guess what their group members deemed the appropriate contribution for each of three hypothetical members. A participant thus makes three guesses, and each of these guesses may be the same as, or different from, the guesses made by the other two group members. We examine the proportion of these three guesses that were exactly the same between all three members of a group (so the possible values

---

[1] The precise score used to assign ranks is: $c_H - \bar{c}_L + 0.02\,\bar{c} + 0.0001R$, where c_H is the participant's view on the appropriate contribution for the high-return member, $\bar{c}_L$ is the participant's view on the appropriate contribution of the two low-return members on average, $\bar{c}$ is the mean appropriate contribution over all three members, and R is a random number between 0 and 1. The addition of $0.02\,\bar{c}$ makes sure that participants who assign a contribution of 20 to all members obtain slightly higher scores than participants who assign a contribution of 0 to all members. This helps to differentiate between different absolute levels of achieving the equal-contributions rule in the sorting method. The number 0.02 is chosen such that whether contributions are relative to returns or not always has dominance in the sorting mechanism over the absolute level of contributions. The addition of $0.0001R$ is to avoid tied scores.

per group are 0, 1/3, 2/3, and 1). The higher this proportion, the more the members share their normative expectations and hence share a social norm. As with the measure of personal normative views, normative expectations are measured three times: once before the first round, once after the first 10 rounds, and once after the last 10 rounds. This allows us to see if social norms emerge over time and whether this differs between conditions. Just after membership change, we additionally asked participants to report their normative expectations towards their new members, but because these correlated highly with normative expectations towards prior group members (correlation = .72, $p < .001$), we do not analyze them separately.

Group identification

Both after the first and second set of 10 rounds, we ask participants to indicate their agreement on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) to the following six statements:

1. I identify with other members of this group
2. I feel strong ties to this group
3. I am like other members of this group
4. This group is an important reflection of who I am
5. I feel proud to be a member of this group
6. I would like to continue working with this group

These items are commonly used to measure group identification in experiments (Leach et al., 2008; Ouwerkerk et al., 1999) We take the average across the six items as the group identification score. This score has very high reliability, as indicated by a Cronbach's alpha of .93. The change in the score before and after membership change is compared between conditions to see how the influence of membership change on group identification depends on normative disagreement between incumbents and newcomers.

Post-experiment measures

After the experiment, participants were asked to provide information on background characteristics such as age, sex, and nationality, as well as some other measures such as their social value orientation. These post-experiment measures are not analyzed in this paper. They are described in detail in the pre-registration of the experiment: osf.io/gy8st.

Results

Normative disagreement. We first examine whether the level of normative disagreement differs by experimental treatment and membership change as intended, i.e., whether our manipulation was successful. Recall that all participants provided their normative view on the appropriate contribution that high-return members and low-return members should make. To measure group-level disagreement, we rank participants within each group based on their ratio of appropriate contributions by high-return members to low-return members. Recall that a ratio of 1 means support for the equal-contributions norm and a ratio of 2 means support for the equal-earnings norm. The level of disagreement of each group is measured by subtracting the ratio supported by the lowest-ranked participant from the ratio supported by the highest-ranked participant. Figure 3a shows the average level of disagreement by experimental condition at three moments in the experiment: the start of the experiment, just after membership change, and at the end of the experiment. We see that, consistent with our intended manipulation, normative disagreements are considerably larger before membership change in condition disagreement-agreement than in condition agreement-disagreement (Mann-Whitney ranksum test, $p < .001$) while the reverse is true after membership change (Mann-Whitney ranksum test, $p < .001$). Thus, in condition disagreement-agreement, membership change leads to a decrease in normative disagreements, whereas in condition agreement-disagreement, membership change leads to an increase in normative disagreements. This can also be seen in the supplementary material, Figure S1, where we show the normative views and disagreement for each group separately.

Social norm emergence. We next turn to the emergence of social norms. Recall that we look at the similarity in normative expectations to examine social norms. Figure 3b presents for both conditions and all three measurement moments the proportion of normative expectations that were the same in all three members per group. As can be seen, condition agreement-disagreement starts with slightly more similarity in normative expectations than condition disagreement-agreement, which is a consequence of sorting for similar normative views (the correlation between normative views and expectations is .64). After the first 10 rounds, the similarity in normative expectations has increased considerably in both conditions (Mann-Whitney ranksum test, $p < .001$ for both conditions), and there is no longer a difference between conditions. This suggests that initial levels of normative disagreement in newly formed groups do not impede the emergence of social norms. However, normative disagreement between

incumbents and newcomers after membership change does seem to matter for the emergence of social norms. The similarity in normative expectations increases when there is less normative disagreement between newcomers and incumbents as in the disagreement-agreement condition. The similarity in normative expectations remains roughly constant when normative disagreement is high as in the agreement-disagreement condition. In other words, groups in which the newcomer agrees with the incumbents develop stronger social norms (Mann-Whitney ranksum test, $p < .001$).

**Figure 3a-b. Normative disagreement and social norm emergence**
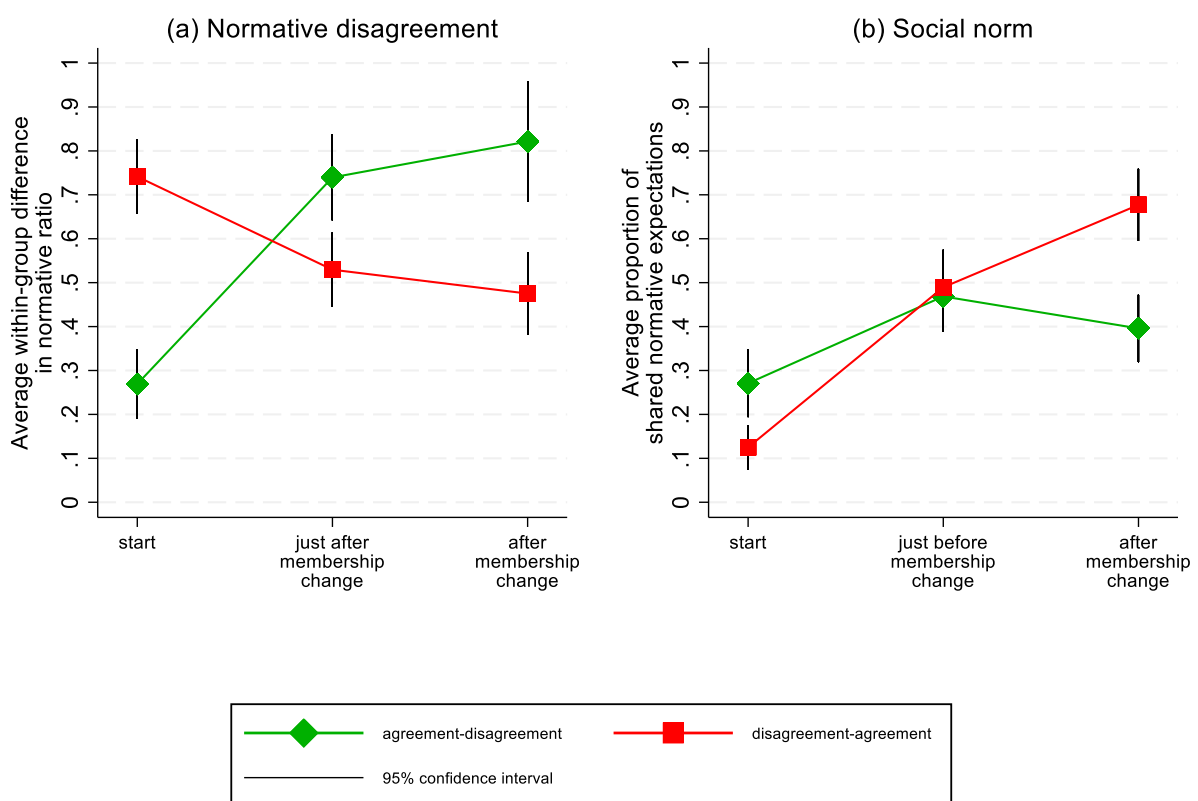
Punishment. Costly peer-punishment is commonly perceived as an instrument of norm enforcement. If someone violates a norm, the person is punished. In this interpretation, more punishment is thus indicative of more norm violations. We now examine if normative disagreement between incumbents and newcomers affects the amount of punishment received, and whether punishment differs between incumbents and newcomers. In Figure 4a, we show the average punishment points sent to each group member per condition and round in groups consisting of newcomers and incumbents. We find that in the early interactions between newcomers and incumbents, punishment levels are considerably higher when there is normative

19

disagreement. In the first three rounds after membership change, the punishment level is about twice as high when newcomers and incumbents disagree instead of agree about how much to contribute (0.61 vs 0.30 punishment points sent to each group member per round, $p < .001$). However, the punishment level decreases in both conditions over time, as does the difference between the conditions. In the last three rounds, we no longer find a significant difference in punishment levels by condition. Normative disagreements between incumbents and newcomers thus spark more costly punishment in early interactions, but not in later interactions. This influence of normative disagreement on costly punishment did not occur in newly formed groups (i.e., before membership change, see supplementary material Figure S2a). There is an outlier in terms of punishment in the last round (round 20), which is related to the so-called endgame effect: contributions tend to drop in the last round of the experiment, leading to higher levels of punishment in both conditions.

In Figure 4b, we show the average punishment points received by incumbents and newcomers for different levels of contributions. We see that there are no significant differences between incumbents and newcomers when they contribute medium to large amounts. In this case, punishment is low for both newcomers and incumbents. However, we find that newcomers are more strongly punished for low contributions. That is, newcomers receive about double the amount of punishment points when they contribute $\leq 25\%$ of their endowment compared to incumbents making these contributions (1.58 vs .76 punishment points received from each group member, $p < .001$). Subdividing by condition suggests that this difference is mostly a result of normative disagreements between incumbents and newcomers. We find that this difference is significant when newcomers and incumbents are in disagreement ($p < .001$), but not when they are in normative agreement ($p = .17$). The differences in punishment points assigned to newcomers and incumbents also appear when examining not their absolute contributions, but how much they deviate from the average contribution in the group (controlling for return-rate differences). Newcomers are more strongly punished than incumbents for deviating from the average contribution, see supplementary material Figure S2b.

Figure 4a-b. Punishment
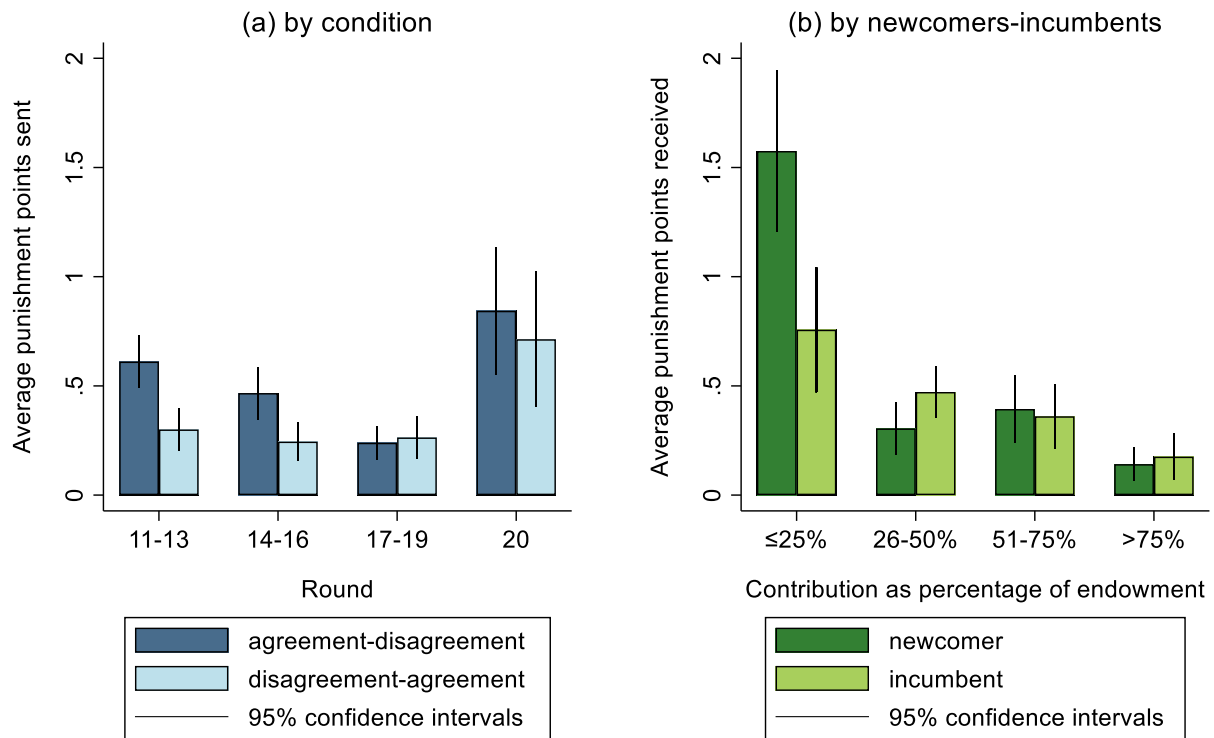
Note: we show the number of punishment points sent by condition and round in panel (a) and the number of punishment points received by newcomer-incumbent division and contribution amount in panel (b). The values in both panels are based on dyads, i.e., punishments sent to, and received from, individual group members. Because each participant has two co-members, the punishment levels will be twice as large in practice. The number of observations totals to 3840 for both panels (192 participants * 10 rounds * 2 co-members to punish or receive punishment from). In panel (b), the number of observations is 428 for contributions ≤25%, 896 for contributions 26-50%, 698 for contributions 51-75%, and 1818 for contributions >75%.

Group identification. We next examine how membership change affects feelings of group identity and how this depends on the level of normative disagreement between newcomers and incumbents. In Figure 5a, we present the participants' group identification by condition before the membership change (measured after round 10) and after the membership change (measured after round 20). We see that while there is no difference between conditions in group identification before membership change, there is a significant difference between conditions after membership change. Feelings of group identification are about 40% higher after membership change in condition disagreement-agreement than in condition agreement-disagreement (Mann-Whitney ranksum test, $p < .001$). We see that the arrival of a newcomer per se does not undermine group identification, it depends on whether the newcomer disagrees with the incumbents. Group identification is lower when newcomers and incumbents disagree (this difference between conditions also holds for each of the six group identification items individually, see supplementary material Figure S3). The differences between experimental

conditions appear for both newcomers and incumbents, but newcomers report on average about 20% less group identification than incumbents (Mann-Whitney ranksum test, *p* = .04).

Figure 5a-b. Group identification and contribution by experimental condition
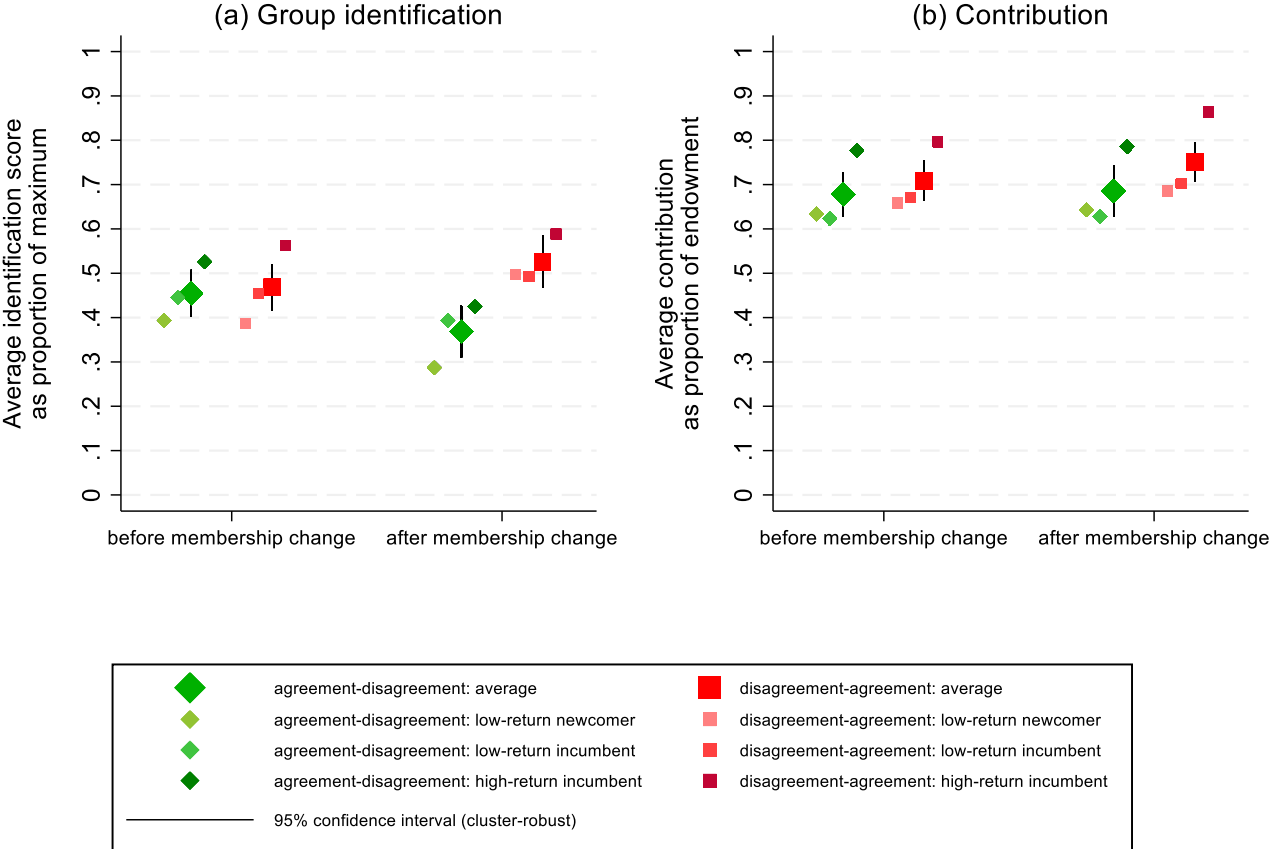
Contributions. To examine how contribution levels change after membership change, we show in Figure 5b the contribution levels by experimental condition both before and after membership change. We find no significant differences in the contribution levels between condition agreement-disagreement and disagreement-agreement, neither before membership change (proportion of endowment contributed of .68 vs .71, cluster-robust *p* = .37) nor after membership change (proportion of endowment contributed of .69 vs .75, cluster-robust *p* = .08). We use population-averaged regression models, which account for repeated measures obtained from the same participant or group, to statistically test the hypothesis that contribution levels decrease more after membership change in condition agreement-disagreement than in condition disagreement-agreement. We take the contribution decision as the dependent variable and as independent variable a factor indicating whether the decision was made in (1) condition agreement-disagreement before membership change, (2) condition disagreement-agreement

22

before membership change, (3) condition agreement-disagreement after membership change, or (4) condition disagreement-agreement after membership change. This allows us to estimate whether the change in contribution levels before and after membership change differs significantly by experimental condition. Across six models, we vary whether the contribution level is on the individual-level or group-level, and whether we include all rounds, only the first rounds (1 and 11), or only the last rounds (10 and 20) as observations. Regardless of which model is used, we find that the change in contribution levels before and after membership change does not differ significantly between conditions according to conventional standards ($p < .05$, with Bonferroni adjustment for multiple comparisons), see supplementary material Table S1. The contribution levels per round can be found in supplementary material Figure S4 and also indicate no difference between conditions. We thus find no support for the hypothesis that the change in contribution levels as a result of membership replacement is more negative in condition agreement-disagreement than in condition disagreement-agreement. In other words, normative disagreements between incumbents and newcomers do not harm cooperation in terms of contributions to the common good.

To sum up the results so far, we find that disagreements between incumbents and newcomers negatively affect social norm emergence and group identification, but not contributions. This suggests that social norms and group identification do not strongly affect contributions. We did not directly manipulate social norm emergence and group identification, making it difficult to get a causal estimate of their effects on contributions. However, since our random assignment to experimental conditions did affect social norm emergence and group identification, we can use the experimental condition as instrumental variable for these two outcomes to enhance our causal inference of their effects on contributions (MacKinnon & Pirlott, 2015; Sobel, 2008). Using the experimental condition as instrumental variable for social norm emergence and group identification, we find little evidence of social norm emergence and group identification affecting contributions ($p = .08$ and $p = .05$ respectively, see Table S2 in the supplementary material).

Predicting contribution levels. To examine what alternatively predicts contribution levels after membership change, and how that differs between newcomers and incumbents, we conduct exploratory analyses. These exploratory analyses can be found in supplementary material S5. Here we explain the main findings. We find that two variables, namely one's own normative view and the contribution of one's group members, explain about half of the variation in

contribution levels. However, the relative influence of these two variables is different for newcomers and incumbents. Newcomers contribute mostly in line with the contributions of incumbents, and only to a small extent with their own views. Incumbents also contribute in line with the contributions of their group members, but to a lesser extent than newcomers. Compared to newcomers, incumbents contribute more in line with their own normative views. We furthermore find that the influence of others' contributions is driven to a large extent by punishment, which holds both for incumbents and newcomers. When moving to the next decision round, participants tend to contribute the same amount if they did not get punished in the prior round. But if they did get punished, they increase their contribution. The stronger the punishment, the more they move away from their prior contribution behavior in favor of higher contributions. With high levels of received punishment (about 10 punishment points), one's prior contribution is no longer related to one's subsequent contribution. However, if participants get punished when contributing large amounts, they do not change their behavior as much.

Because newcomers act more in line with the contribution of incumbents, and the contribution of incumbents is related to their normative views, newcomers end up conforming to the incumbents' normative views instead of their own. The relative influence of the newcomer's own normative views and that of the incumbents is best shown in condition agreement-disagreement, because that is where the newcomers and incumbents hold different normative views and thus have to decide what view to conform to. In Figure 6, we show for newcomers (a) and incumbents (b) how their contributions correlate with their own normative view and the normative view of their group members. In Figure 6a, we see that while newcomers start by contributing according to their own normative view, they very quickly disregard their own view and contribute almost fully according to the view of the incumbents. The correlation between their contribution and own normative view is .66 in round 11 and drops to only .24 one round later. The correlation between their contribution and their group members' normative view is only .09 in round 11 and increases to .57 in the subsequent round. Newcomers thus contribute according to their group members' normative view already from the second interaction after they have entered the group. In contrast, we see in Figure 6b that incumbents contribute more in line with their own normative view. The correlation between their contribution and own normative view hovers around .75 throughout almost all rounds, whereas the correlation between their contribution and their group members' normative view hovers around .40. These findings suggest that when there is disagreement between incumbents and newcomers on how to contribute to the public good, the newcomers largely concede to incumbents.

Although newcomers change their behavior to act in line with the incumbents' normative views, newcomers do not change their own normative views. In supplementary material Figure S7, we show that both newcomers and incumbents hold relatively stable normative views. The average temporal change in what they view to be appropriate contributions (when comparing views before and after the 10 rounds they interact together) is about 2 to 3 contribution points out of 20 for both incumbents and newcomers. Thus, while newcomers may change their behavior to adapt to incumbents, their normative views do not change much. This means that groups end up with inequality in the extent to which normative views are realised, with incumbents being more likely to realise their views than newcomers.

Figure 6a-b. Correlation between contribution and own or others' normative views
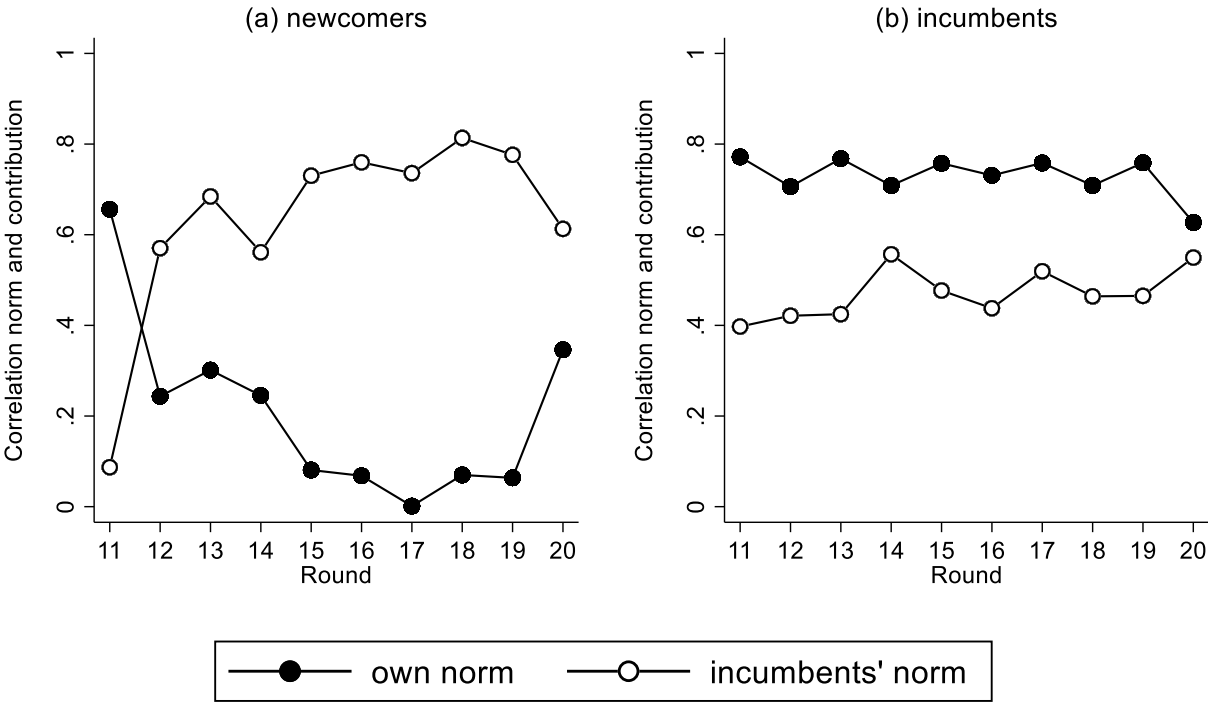
Note: All participants provided their normative view on the appropriate contribution that a high-return member should make and the appropriate contribution that each of the low-return members should make. We examine if participants contribute in line with their normative view on how much they themselves should contribute (i.e., if they have a low-return, how much they think low-return members should contribute; if they have a high-return, how much high-return members should contribute) or in line with how much their group members think they should contribute. The results are separated by newcomers and incumbents and shown for condition agreement-disagreement, as this is the condition where the newcomers and incumbents hold different normative views and thus have to decide what view to conform to. We use the normative views as measured just before the membership change (round 10).

In each group, there are 2 incumbents and 1 newcomer. When comparing newcomers and incumbents, we are thus also comparing a minority with the majority. It is possible that this aspect alone, being in the minority or majority group, already generates the results found in

Figure 6. To examine this possibility, we can look at groups *before* membership change in condition disagreement-agreement. Here, there are no incumbent-newcomer divisions yet and one member per group disagrees with the two other members, i.e., one member holds a minority normative view and the other two members hold a majority normative view. In supplementary material Figure S6, we show that we find no clear differences between minority and majority participants in how contribution decisions correlate with their own normative view and their group members' normative views. Thus, newcomers conceding to the normative views of incumbents seems not just to be the result of the newcomers being a minority.

Discussion

Groups often rely on shared norms to achieve cooperation. These shared norms may be challenged when new members enter the group and old members leave. If newcomers hold normative views that are incompatible with the normative views of incumbents, a normative disagreement ensues that may harm cooperation. Using the public goods game paradigm, we studied experimentally whether the impact of membership change on cooperation depends on the level of normative disagreement between incumbents and newcomers. While we find that normative disagreement negatively affects newcomer-incumbent relations in terms of group identification and social norm emergence, we find that contributions to the common good are not impacted. Newcomers and incumbents can thus cooperate for the provision of common goods even in the presence of normative disagreement and the associated lower levels of group identification and social norm emergence. Our findings suggest an important role for norm enforcement in sustaining cooperation in groups consisting of newcomers and incumbents. Norm enforcement was possible in our experiment via peer punishment. Low contributing participants respond strongly to punishment by increasing their subsequent contributions, and newcomers in particular are punished strongly for making low contributions and deviations from the incumbents' contribution norm. The result is that newcomers mostly act in line with the incumbents' norm.

These results paint a nuanced picture of the effects of normative disagreements between newcomers and incumbents. While it may be reassuring that contribution levels are unaffected, the newcomer-incumbent relations, as measured by group identification and social norm emergence, are negatively impacted by normative disagreements. And while cooperation is achieved, it is mostly at a level supported by the incumbents and not the newcomers. There is thus a newcomer-incumbent inequality in the extent to which normative views are realised.

Similarly, the punishment results indicate a newcomer-incumbent inequality in terms of how strongly low contributors are punished. We can thus not unequivocally conclude whether normative disagreements are harmful; it depends on which outcome measures are considered. Focusing only on the achieved contribution level as outcome may lead to a misrepresentation of how group members are doing in terms of subjective experiences.

The prediction that cooperation for public good provision is harmed when newcomers and incumbents do not share the same norms is relatively common, especially when considering cooperation between immigrants and natives (Collier, 2013; Habyarimana et al., 2009; Ostrom, 2000). It might therefore come as a surprise that we find no support for it in our experiment. However, a recent literature review on cooperation between natives and immigrants suggests that prior estimates for the effect of normative differences between immigrants and natives on public good provision are often confounded with other aspects of interethnic groups, such as poverty and political instability (Baldassarri & Abascal, 2020). To get a causal estimate on the effect of normative differences between newcomers and incumbents on group cooperation, experimental manipulation such as in our study is helpful to isolate the effect of normative differences from potential confounders. Our study suggests that normative disagreements alone do not predict whether newcomers and incumbents cooperate towards public good provision.

Indeed, an emerging literature challenges the commonly held view that the perceived threats of immigration are of an economic nature, and instead suggests that the perceived threats are of a cultural nature (Hainmueller & Hopkins, 2014). Our findings that disagreements between newcomers and incumbents do not harm contributions to the public good but do harm group identification and impede social norm emergence seem to corroborate this view. That is, we found no economic impacts in terms of contribution to the public good, but did find impacts in terms of group identification and social norm emergence. Our findings suggest that the absence of an impact on contributions is partly related to norm enforcement. Participants that clearly deviate from prosocial contribution norms are punished and react by increasing their contribution. This norm enforcement is especially directed at newcomers, who are punished twice as much as incumbents for contributing low amounts. A recent field experiment on norm enforcement in Germany found a similar pattern; natives impose norms on immigrants considerably more so than the other way around (Winter & Zhang, 2018).

It is interesting to see that the negative effects of normative disagreement on social norm emergence only appeared after membership change, i.e., when the disagreements were between incumbents and newcomers. In newly formed groups (i.e., before membership change), social norms strengthened over time regardless of the level of normative disagreement. This strengthening over time stopped after membership change when there was normative disagreement between new and incumbents, whereas it continued when there was no normative disagreement between newcomers and incumbents. Likewise, normative disagreements did not negatively impact group identification before membership change, whereas group identification after membership change was lower with higher levels of normative disagreement. This is consistent with the idea that newcomer-incumbent division, like intergroup division, threatens shared identities and norms if the newcomers do not share the incumbents' normative views. Perhaps early group interactions before membership change shape the participant's expectations on how groups ought to behave, and these expectations are then exceeded or disappointed after membership change depending on whether normative disagreement increases or decreases. Regardless of what exactly drives these differences before and after membership change, our findings suggest that we cannot expect the results found in newly formed groups to generalize to reshaped groups consisting of newcomers and incumbents.

As with any experimental study, our conclusions might depend on design choices. We tried to stay close to the prior studies on normative disagreements in PGGs in our design. Future research can assess to what extent our findings also hold under different designs. We provide a few suggestions. We studied normative disagreements between the norms of equal contributions and equal earnings, both of which are common and important norms in heterogeneous groups (Nikiforakis et al., 2012; Reuben & Riedl, 2013). Future research can study whether disagreements differently impact cooperation when other norms are involved. For example, in some contexts different norms may produce different benefits to different subgroups. If some norms clearly favor the incumbents while other norms favor the newcomers, normative disagreements might be further accentuated by ingroup favoritism. Moreover, we studied groups of three in which one incumbent is replaced by one newcomer. Future research can vary the number of newcomers and their relative share. Normative disagreements might be more harmful when the number and share of newcomers is larger, as this increases the chance for separate ingroup-outgroup subdivision and the newcomers' power to influence norms (Pettigrew, 1991). In addition, our results suggest that newcomers concede to the views of incumbents as a result of receiving punishment when sticking to their own contribution

standards. Perhaps normative disagreements will be more problematic for cooperation if such punishment is not possible, which could be studied with the public goods game without punishment. Finally, future research could examine newcomer-incumbent cooperation using natural rather than minimal groups, for example by letting immigrants and natives play the public goods game together (Drouvelis et al., 2019). Although natural identities are conflated with potential confounders (e.g., differences in status and income), which make it more difficult to isolate theoretical mechanisms, they can enhance the external validity and thereby complement research using minimal groups.

We conclude that the absence of conflict in terms of cooperation failure does not imply that newcomer-incumbent relations are harmonious. Our results suggest that a fuller understanding of newcomer-incumbent relations is achieved when multiple dimensions are measured simultaneously, e.g., behavior (in our case contributions and punishment), subjective experiences (in our case group identification), and normative aspects (in our case normative views and expectations). As we showed, the results between these different dimensions need not be in line with each other. While this makes it more difficult to draw clear conclusions, we think it is important to allow for a nuanced view on intergroup relations. High levels of contributions to the common good are not always unequivocally good or bad; it may depend on how group members think about it. We showed that normative disagreements between newcomers and incumbents do not lead to conflict in terms of cooperation failure. Yet, they do seem to negatively impact group identification and lead to contribution levels that predominantly reflect the normative views of incumbents rather than newcomers.

References

Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2020). *Predicting social tipping and norm change in controlled experiments.* (NBER Working Paper No. 27310). https://doi.org/10.3386/w27310

Arechar, A., Kouchaki, M., & Rand, D. G. (2018). Examining spillovers between long and short repeated prisoner's dilemma games played in the laboratory. *Games*, *9*(1), 5. https://doi.org/10.2139/ssrn.2979639

Arrow, H., & McGrath, J. E. (1993). Membership matters: How member change and continuity affect small group structure, process, and performance. *Small Group Research*, *24*(3), 334–361.

Baldassarri, D., & Abascal, M. (2020). Diversity and prosocial behavior. *Science*, *369*(6508),

1183–1187. https://doi.org/10.1126/SCIENCE.ABB2432

Balliet, D., Wu, J., & Dreu, C. K. W. De. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, *140*(6), 1556–1581. https://doi.org/http://dx.doi.org/10.1037/a0037737

Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M., & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization : A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology*, *92*(3), 707–721. https://doi.org/10.1037/0021-9010.92.3.707

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review*, *96*(2), 217–221. https://doi.org/10.1257/000282806777212594

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*(7105), 912–915. https://doi.org/10.1038/nature04981

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C., Lindemans, J. W., & Jiang, T. (2014). A structured approach to a diagnostic of collective practices. *Frontiers in Psychology*, *5*(DEC), 1–13. https://doi.org/10.3389/fpsyg.2014.01418

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict : A review of theories and measures. *Journal of Economic Behavior and Organization*, *178*, 947–962. https://doi.org/10.1016/j.jebo.2018.01.020

Carpenter, J., & Matthews, P. H. (2009). What norms trigger punishment? *Experimental Economics*, *12*(3), 272–288. https://doi.org/10.1007/s10683-009-9214-z

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. https://doi.org/10.1007/s10683-010-9257-1

Collier, P. (2013). The social consequences. In *Exodus: How migration is changing our world* (pp. 57–110). Oxford University Press.

De Cremer, D., & van Knippenberg, D. (2003). Cooperation with leaders in social dilemmas: On the effects of procedural fairness and outcome favorability in structural cooperation. *Organizational Behavior and Human Decision Processes*, *91*, 1–11. https://doi.org/10.1016/S0749-5978(02)00539-3

De Dreu, C. K. W., Gross, J., Fariña, A., & Ma, Y. (2020). Group Cooperation, Carrying-Capacity Stress, and Intergroup Conflict. *Trends in Cognitive Sciences*, *24*(9), 760–776.

https://doi.org/10.1016/j.tics.2020.06.005

Diekmann, A., & Przepiorka, W. (2016). "Take one for the team!" individual heterogeneity and the emergence of latent norms in a volunteer's dilemma. *Social Forces*, *94*(3), 1309–1333. https://doi.org/10.1093/sf/sov107

Dorrough, A. R., Glöckner, A., Hellmann, D. M., & Ebert, I. (2015). The development of ingroup favoritism in repeated social dilemmas. *Frontiers in Psychology*, *6*(APR). https://doi.org/10.3389/fpsyg.2015.00476

Drouvelis, M., Malaeb, B., Vlassopoulos, M., & Wahba, J. (2019). *Cooperation in a fragmented society : Experimental evidence on Syrian refugees and natives in Lebanon* (IZA Discussion Papers No. 12858).

Duffy, J., & Lafky, J. (2016). Birth, death and public good provision. *Experimental Economics*, *19*(2), 317–341. https://doi.org/10.1007/s10683-015-9439-y

Duffy, J., & Lafky, J. (2019). *Social conformity under evolving private preferences*. Working Paper. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3226057

Duffy, J., & Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, *66*(2), 785–812. https://doi.org/10.1016/j.geb.2008.07.003

Dur, R., & Sol, J. (2010). Social interaction , co-worker altruism , and incentives. *Games and Economic Behavior*, *69*(2), 293–301. https://doi.org/10.1016/j.geb.2009.10.013

Engl, F., Riedl, A., & Weber, R. A. (2017). *Spillover effects of institutions on cooperative behavior, preferences and beliefs*. (CESifo Working Paper No. 6504).

Fehr, E, & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, Ernst, & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fehr, Ernst, & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, *2*(7), 458–468. https://doi.org/10.1038/s41562-018-0385-5

Fréchette, G. R., & Yuksel, S. (2017). Infinitely repeated games in the laboratory: four perspectives on discounting and random termination. *Experimental Economics*, *20*(2), 279–308. https://doi.org/10.1007/s10683-016-9494-z

Gangadharan, L., Nikiforakis, N., & Villeval, M. C. (2017). Normative conflict and the limits of self-governance in heterogeneous populations. *European Economic Review*, *100*, 143–156. https://doi.org/10.1016/j.euroecorev.2017.07.013

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with

ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Grund, C., Harbring, C., & Thommes, K. (2015). Public good provision in blended groups of partners and strangers. *Economics Letters*, *134*, 41–44. https://doi.org/10.1016/j.econlet.2015.06.008

Guala, F., & Mittone, L. (2010). How history and convention create norms: An experimental study. *Journal of Economic Psychology*, *31*(4), 749–756. https://doi.org/10.1016/j.joep.2010.05.009

Guido, A., Robbett, A., & Romaniuc, R. (2019). Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence. *Journal of Economic Behavior and Organization*, *159*, 192–209. https://doi.org/10.1016/j.jebo.2019.02.009

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2009). *Coethnicity: diversity and the dilemmas of collective action*. Russell Sage Foundation.

Hainmueller, J., & Hopkins, D. J. (2014). Public attitudes toward immigration. *Annual Review of Political Science*, *17*, 225–249. https://doi.org/10.1146/annurev-polisci-102512-194818

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus : Behavioral experiments in 15 small-scale societies. *American Economic Review*, *91*(2), 73–78. https://doi.org/10.1257/aer.91.2.73

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808

Jetten, J., Spears, R., & Manstead, A. S. R. (1996). Intergroup norms and intergroup discrimination : Distinctive self-categorization and social identity effects. *Journal of Personality and Social Psychology*, *71*(6), 1222–1233. https://doi.org/10.1037/0022-3514.71.6.1222

Kingsley, D. C. (2016). Endowment heterogeneity and peer punishment in a public good experiment : Cooperation and normative conflict. *Journal of Behavioral and Experimental Economics*, *60*, 49–61. https://doi.org/10.1016/j.socec.2015.12.002

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495–524. https://doi.org/10.1111/jeea.12006

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, *90*, 375–402. https://doi.org/10.1016/j.euroecorev.2015.11.011

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-Level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Levine, J. M., & Moreland, R. L. (1994). Group socialization: Theory and research. *European Review of Social Psychology, 5*(1), 305–336. https://doi.org/10.1080/14792779543000093

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, *19*(1), 30–43. https://doi.org/10.1177/1088868314542878

McCarter, M. W., & Sheremeta, R. M. (2013). You can't put old wine in new bottles: The effect of newcomers on coordination in groups. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0055058

Moreland, R. L. (1985). Social categorization and the assimilation of "new" group members. *Journal of Personality and Social Psychology*, *48*(5), 1173–1190. https://doi.org/10.1037/0022-3514.48.5.1173

Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, *96*(9–10), 797–807. https://doi.org/10.1016/j.jpubeco.2012.05.014

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, *14*(3), 137–158. https://doi.org/10.1257/jep.14.3.137

Otten, K., Buskens, V., Przepiorka, W., & Ellemers, N. (2020). Heterogeneous groups cooperate in public good problems despite normative disagreements about individual contribution levels. *Scientific Reports*, *16702*, 1–12. https://doi.org/10.1038/s41598-020-73314-7

Ouwerkerk, J. W., Ellemers, N., & Kortekaas, P. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology*, *29*(2–3), 371–389. https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/3<371::AID-EJSP932>3.0.CO;2-U

Pettigrew, T. F. (1991). Normative theory in intergroup relations : Explaining both harmony and conflict. *Psychology and Developing Societies*, *3*(1), 3–16. https://doi.org/10.1177/097133369100300102

Peysakhovich, A., & Rand, D. G. (2013). Habits of virtue: Creating norms of cooperation and

defection in the laboratory. *Management Science*, *62*(3), 631–647.
https://doi.org/10.1021/la00030a006

Ranehill, E., Schneider, F., & Weber, R. A. (2014). *Growing groups, cooperation, and the rate of entry* (CESifo Working Paper No. 4719). https://ssrn.com/abstract=2425583

Rauhut, H., & Winter, F. (2017). Types of normative conflicts and the effectiveness of punishment. In W. Przepiorka & B. Jann (Eds.), *Social dilemmas, institutions, and the evolution of cooperation* (Issue January, pp. 239–256).

Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, *77*(1), 122–137.
https://doi.org/10.1016/j.geb.2012.10.001

Reuben, E., Riedl, A., & Bernard, M. (2015). *Fairness and coordination: The role of fairness principles in coordination failure and success*. Working Paper. Available at:
https://sites.google.com/site/markbernard1984/research-papers

Rink, F., Kane, A. A., Ellemers, N., & van der Vegt, G. (2013). Team receptivity to newcomers: Five decades of evidence and future research themes. *Academy of Management Annals*, *7*(1), 247–293. https://doi.org/10.1080/19416520.2013.766405

Sanders, K. (2009). Cooperative behaviours in organizations. In R. Morrison & S. Wright (Eds.), *Friends and Enemies in Organizations: A Work Psychology Perspective* (pp. 101–121). London: Palgrave Macmillan. https://doi.org/10.1057/9780230248359

Smerdon, D., Offerman, T., & Gneezy, U. (2019). 'Everybody's doing it': on the persistence of bad social norms. *Experimental Economics*, *0123456789*.
https://doi.org/10.1007/s10683-019-09616-z

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics*, *33*(2), 230–251.
https://doi.org/10.3102/1076998607307239

Sonnemans, J., Schram, A., & Offerman, T. (1999). Strategic behavior in public good games: When partners drift apart. *Economics Letters*, *62*(1), 35–41.
https://doi.org/10.1016/s0165-1765(98)00203-1

Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition*, *167*, 212–254.
https://doi.org/10.1016/j.cognition.2017.01.017

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations. Key readings in*

social psychology (pp. 7–24). Nelson-Hall.

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37–40. https://doi.org/10.1016/j.econlet.2018.06.022

Thravalou, E., Martinovic, B., & Verkuyten, M. (2020). Humanitarian assistance and permanent settlement of asylum seekers in Greece : The role of sympathy, perceived threat, and perceived contribution. *International Migration Review*, 1–27. https://doi.org/10.1177/0197918320949826

Titlestad, K., Snijders, T. A. B., Durrheim, K., Quayle, M., & Postmes, T. (2019). The dynamic emergence of cooperative norms in a social dilemma. *Journal of Experimental Social Psychology*, *84*(103799). https://doi.org/10.1016/j.jesp.2019.03.010

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory.* Oxford: Basil Blackwell. https://doi.org/10.2307/2073157

van Gerwen, N., Buskens, V., & van der Lippe, T. (2018). Individual training and employees' cooperative behavior: Evidence from a contextualized laboratory experiment. *Rationality and Society*, *30*(4), 432–462. https://doi.org/10.1177/1043463118771428

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, *40*(1), 145–180. https://doi.org/10.2307/2393703

Winter, F., Rauhut, H., & Helbing, D. (2012). How norms can generate conflict: An experiment on the failure of cooperative micro-motives on the macro-level. *Social Forces*, *90*(3), 919–948. https://doi.org/10.1093/sf/sor028

Winter, F., & Zhang, N. (2018). Social norm enforcement in ethnically diverse communities. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2722–2727. https://doi.org/10.1073/pnas.1718309115