



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Cooperation between newcomers and incumbents: The role of normative disagreements

Kasper Otten^{a,*}, Vincent Buskens^a, Wojtek Przepiorka^a, Naomi Ellemers^b

^a Utrecht University, Department of Sociology, the Netherlands

^b Utrecht University, Department of Psychology, the Netherlands

ARTICLE INFO

Keyword:

Newcomers
Cooperation
Normative disagreement
Social norms
Punishment
Group identification

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 and lowers feelings of group identification. Contrary to expectations, normative disagreement does not affect cooperation negatively. Instead, participants adjust their behavior to each other's standards, using punishment for norm enforcement. This punishment is especially directed at low-contributing newcomers, leading them to conform to the incumbents' higher contribution standards.

1. 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, Buskens, & van der Lippe, 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

* Corresponding author at: Padualaan 14, 3584 CH Utrecht, the Netherlands.

E-mail address: k.d.otten@uu.nl (K. Otten).

<https://doi.org/10.1016/j.joep.2021.102448>

Received 28 January 2021; Received in revised form 19 July 2021; Accepted 20 September 2021

Available online 28 September 2021

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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, Humphreys, Posner, & Weinstein, 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). In social dilemmas, individual and collective interests are at odds. Cooperation in social dilemmas thus requires that people forego their individual interest to act in line with the collective interest. For example, taking measures to reduce one's carbon footprint are often individually costly but benefit society. When one's individual interest is aligned with cooperation, no social dilemma exists, and cooperation is more easily achieved. However, we focus on social dilemmas, which means that cooperation is arguably more fragile and vulnerable to normative disagreement. Most social dilemma research is directed either at intragroup cooperation (i.e., cooperation between members of the same group) or intergroup cooperation (i.e., cooperation between members of different groups). A main finding is that intragroup cooperation can be sustained via the development and enforcement of contribution norms, i.e., norms on how much to contribute to the collective good (Fehr & Schurtenberger, 2018). Although these norms promote intragroup cooperation, they can impede intergroup cooperation (De Dreu, Gross, Fariña, & Ma, 2020). Norms are typically group-specific and favor the ingroup over the outgroup (Bernhard, Fischbacher, & Fehr, 2006; Titlestad, Snijders, Durrheim, Quayle, & Postmes, 2019). This can lead to conflict when different groups have to cooperate together but each group wants to stick to their own norm (Jetten, Spears, & Manstead, 1996). Indeed, there is ample research showing that intergroup relations are often characterized by conflict rather than cooperation (Balliet, Wu, & Dreu, 2014; Böhm, Rusch, & Baron, 2020).

While research using social dilemma games has advanced our knowledge about intragroup and intergroup cooperation, we know little about cooperation in groups where newcomers enter and have to cooperate with incumbents to contribute to the common good. We regard newcomer-incumbent relations as an intermediate case between intragroup and intergroup relations. Studying these intermediate cases is important because newcomers, although situated within the incumbent-group, are often seen as outsiders by the incumbents (Rink, Kane, Ellemers, & van der Vegt, 2013) and also themselves do not readily identify with the incumbents (Moreland, 1985). The change in group composition brought about by newcomer entry and the resulting newcomer-incumbent relations may have consequences for contributions to the common good.

Stability in group composition is commonly theorized to promote contributions to the common good via shared social norms (Fehr & Schurtenberger, 2018). 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, Nikiforakis, & Villeval, 2017; Henrich et al., 2001). When members migrate between groups with different and incompatible norms, normative disagreement ensues. When members instead migrate between groups that hold similar norms, the arrival of newcomers does not increase normative disagreement.

According to normative conflict theory (Rauhut & Winter, 2017; Winter, Rauhut, & Helbing, 2012), groups are able to cooperate by contributing high amounts for the collective good if their members agree on normative views and can enforce these views, for example by the punishment of norm violations. However, if group members disagree on normative views, conflict is expressed in both low contribution levels and high punishment levels. When members differ in their normative views and contribute according to their own view, they observe that their view is not adhered to by others. The expected consequence is that members who feel others are not contributing enough reduce their own contribution and/or punish those they disapprove of. While punishment typically helps to promote cooperation if members agree on their normative views, it is predicted to be harmful when members disagree on normative views. Members who receive punishment while behaving according to their own normative view deem the punishment unjustified, and retaliate by reducing their contribution and by counter-punishment, leading to lower contribution levels and higher punishment levels.

Hence, if a newcomer brings in a normative view that is at odds with the normative views of the incumbents, a normative disagreement ensues that is predicted to lead to conflict in terms of cooperation failure. If the newcomer's normative view does not conflict with those of the incumbents, there is normative agreement, and cooperation is expected to be sustained. 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. Some of these studies suggest a positive effect of newcomers on contributions to the common good (e.g., Duffy & Laffky, 2016; Sonnemans, Schram, & Offerman, 1999), whereas other studies report a negative effect (e.g., Grund, Harbring, & Thommes, 2015; Ranehill, Schneider, & Weber, 2014). This prior literature has not yet examined the role of normative disagreement between newcomers and incumbents in explaining the level of cooperation. Using a public goods game, we sort participants into groups of three and let them make contribution decisions in two sets of rounds. After the first set, we replace one member per group for a member from another group, such that each group consists of one newcomer and two incumbents. We manipulate whether newcomers and incumbents agree or disagree on normative views and examine if this influences cooperation in the second set of rounds.

Although newcomers do not always come from a salient outgroup, there are many situations in which they do. For example, work organizations may hire employees that previously worked at a competitor to obtain inside information from that competitor, players in professional sports are regularly switched between competing teams, and immigration also often involves ingroup-outgroup

considerations in social interactions and conflict (Hainmueller & Hopkins, 2014). We focus on a situation where a newcomer comes from a relatively salient outgroup.¹

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).

2. Previous research

2.1. Membership changes in public good games

One of the earliest experiments on membership changes in the public goods game was conducted by Sonnemans et al. (1999). The study examined 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 found 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 Laffky (2016) found that periodically replacing old members by new members helps 4-player groups to sustain contributions 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) examined contribution levels in 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 found that the contribution level is lower in groups with more strangers (i.e., more membership changes). Ranehill et al. (2014) studied how the rate of newcomer entry affects contribution levels in growing groups. They found that a higher rate of newcomer entry (i.e., more newcomers at once vs each newcomer entering in separate rounds) negatively affects contribution levels. Salmon and Weber (2017) also find that higher rates of newcomer entry negatively affect contribution levels. In addition, they find that restrictions via entry quotas that limit the number of newcomers or via entry quizzes in which newcomers need to demonstrate sufficient competence can help to reduce this negative effect. Finally, McCarter and Sheremeta (2013) found that newcomers have a negative effect on the effort devoted to group cooperation.²

The discussed studies generally looked at overall effects, i.e., whether newcomers affect contributions to the public good, and produced mixed results. Potential mechanisms that drive newcomer effects have received much less attention. Our study focuses on the role of normative disagreements between incumbents and newcomers. In the next section, we turn to 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.

2.2. Intergroup differentiation in newcomer-incumbent 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, Bodner, Erdogan, Truxillo, & Tucker, 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, Hogg, Oakes, Reicher, & Wetherell, 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

¹ Before the game starts, participants are told that each group receives a color: half the groups a blue color and half the groups an orange color. When switching members across groups, the incumbents are explicitly told that the newcomer comes from a group with a different color, and the newcomer is also explicitly told that the incumbents have a different color from the newcomer's prior group. Such minimal groupings have been shown to make group identities salient and lead to intergroup biases (Lane, 2016).

² There is a related literature where individuals can form groups and thus also have the option to enter and leave groups (endogenous group formation). However, these studies are designed to examine group formation processes rather than effects of membership changes. Because the role of newcomer and incumbent are assumed endogenously in these studies, causal effects of newcomer entry and newcomer status are more difficult to identify. We therefore refrain from reviewing this literature. However, a recent overview can be found in Guido, Robbett, and Romaniuc (2019).

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, Martinovic, & Verkuyten, 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 may 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 briefly review how the effect of normative disagreement on cooperation in public goods games has been studied so far.

2.3. Normative disagreement in public good games

Prior experiments on normative disagreement in public good games typically study disagreement between the commonly supported norms of equal-contributions and equal-earnings (Gangadharan et al., 2017; Nikiforakis, Noussair, & Wilkening, 2012; Reuben & Riedl, 2013). The tension between these norms arises when there are heterogeneous returns of the public good, i.e., when some group members benefit more from the public good than others. The equal-contributions norm prescribes that all members contribute equally, which implies that those who obtain a higher return from the public good also end up earning more. The equal-earnings norm prescribes that members who obtain a higher return from the public good also contribute more such that earnings are equalized. Several studies suggest that groups disagreeing on these two norms are more likely to fail to cooperate in terms of contributing to the public good (Kingsley, 2016; Nikiforakis et al., 2012; Rauhut & Winter, 2017; Winter et al., 2012). In these studies, 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 is hypothesized to work via observing each other's contributions and judging whether these contributions match one's own normative view. 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.

While punishment of low contributors mostly promotes cooperation if members agree on their normative views, it is predicted to be ineffective or harmful when members disagree on normative views. If members feel that their contribution is sufficient and nevertheless are punished, they may refuse to increase their contribution and retaliate against the other members with counter-punishment. Groups with normative disagreement are thus expected to obtain lower contribution levels than groups with normative agreement when punishment is possible. Prior research suggests that without punishment opportunities most groups show a trend to free-riding regardless of any between-group differences in normative disagreement (Reuben & Riedl, 2013).

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, Buskens, Przepiorka, & Ellemers, 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, Nikiforakis, & Siegenthaler, 2021; Duffy & Laffky, 2021; Guala & Mittone, 2010; Smerdon, Offerman, & Gneezy, 2019). Furthermore, participants take the norms they have learned in prior interactions with them when entering new social situations (Engl, Riedl, & Weber, 2017; Peysakhovich & Rand, 2013; Stagnaro, Arechar, & Rand, 2017), although not indefinitely (Arechar, Kouchaki, & Rand, 2018; Duffy & Ochs, 2009). Hence, newcomers are expected to be attached to the norm of their prior group, which may impede cooperation if this norm is different from that of the incumbents. We will test the following hypothesis:

Hypothesis. Normative disagreement between newcomers and incumbents harms cooperation in terms of contributions to the public good.

3. 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. Sample size was determined based on sample sizes in comparable studies (see for example Nikiforakis et al., 2012; Reuben & Riedl, 2013) and desired power. A Mann–Whitney ranksum test with individuals as unit of analysis has a high power (0.96) to detect medium-sized effects (Cohen's $d = 0.5$) given our sample size. We

deem a medium-sized effect of Cohen's $d = 0.5$ reasonable, as this is close to the average effect size found in the social dilemma literature (average effect size in the Cooperation Databank is 0.49; Spadaro et al., 2020). 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 min. 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 & Gächter, 2000).

3.1. 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 $c_i \in \{0, 1, \dots, 20\}$. The part of the endowment that i does not contribute is kept for oneself. The public good benefits each group member j and consists of the sum of the contributions made by all members of the group $\sum_j c_j$. We use a group size of $N = 3$. Each member receives a return m_i per MU 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 $p_{ij} \in \{0, 1, \dots, 10\}$ to each group member $j \neq i$. Each assigned punishment point p_{ij} costs 1 MU to the punisher and each received punishment point p_{ji} reduces the payoff of the punished group member by 3 MU. 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 public good games 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 analyze 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 $m_i = 0.50$ from the public good and one member a higher return of $m_i = 0.75$ (this makes the joint multiplication factor $M = 1.75$). Participants know which members have the low-return 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).

3.2. Normative views and expectations

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 ($m_i = 0.75$) from the public good. The exact same composition of returns is 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 views are measured again after the first 10 rounds of the game (before announcing the membership change), 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.

Every time after participants report their personal normative views, we tell them that their group members were also asked to indicate appropriate contributions for three members in the public goods game. Each participant is then asked to guess the answers submitted by their group members, i.e., to report their normative expectations. To incentivize the normative expectation, participants are informed that one of their guesses will be picked randomly and yield an additional payment of 100 MU ($\sim \text{€}1.40$) 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, Riedl, & Bernard, 2015).

Participants' normative views are used to position these participants 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.³

We regard normative views as attitudes on what is appropriate. Unlike preferences, attitudes cannot be incentivized without running the risk of introducing confounds. Although it is possible to link a participant's normative view to monetary consequences for other unrelated participants (similar to spectator methods), this still runs the risk of confounding by other-regarding and distributional preferences (for example, a competitive spectator may wish to allocate few points to others, even though the spectator does think it is appropriate that others receive more points). For this reason, normative views are often elicited in a non-incentivized way (Bicchieri, 2006; Cubitt, Drouvelis, Gächter, & Kabalin, 2011; Hauge, 2015). However, as it turns out, normative views correlate highly with normative expectations in our experiment (correlation of 0.71, $p < .001$). Hence, in practice, it would not have made much of a difference if we used the incentivized normative expectations for group sorting or the unincentivized normative views. What is more, participants' normative views are also strongly correlated with their incentivized contribution behavior after the measurement. A participant's normative view measured after round 10 correlates at 0.72 with the participant's contribution decision in round 11. This gives additional indication that participants did take the measurement of normative views seriously.

3.3. 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, as shown in Fig. 1. In the first condition (upper part of Fig. 1), we initially sort participants from the same side of the spectrum together, whereas in the second condition (lower part of Fig. 1) we initially sort participants from different sides of the spectrum together. 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. The first 10 rounds of the experiment are analyzed in another study (Otten et al., 2020) and we do not find a difference between conditions in outcomes (contributions, group identification, social norms) at the end of these first 10 rounds (also shown in supplementary material, section S2). In the current study, we focus on the second part of the experiment.

After the 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 consists of one newcomer and two incumbents. We inform the participants of this membership change, and communicate to the incumbents that the newcomer comes from a group with a different color. Likewise, we communicate to the newcomer that the entered group is of a different color than that of the newcomer's prior group. We then let the reshaped groups play a second set of 10 rounds. How members are exchanged differs by condition. In the first condition, we replace a member that had a similar normative view to the two other members by a newcomer that has a dissimilar normative view to the two other members (upper part of Fig. 1). Here, the newcomer is thus in normative disagreement with two incumbents. In the second condition, we replace a member that had a dissimilar normative view to the two other members by a newcomer that has a similar normative view to the two other members (lower part of Fig. 1). Here, the newcomer is thus in normative agreement with the two incumbents. By comparing the contribution levels between both conditions during the second set of 10 rounds, we can test our hypothesis of the negative influence of normative disagreement among newcomers and incumbents on cooperation. 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.

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 Fig. 2.

Note that the variation in normative views among participants is used *within* conditions to sort them into groups. However, the manipulation *between* conditions (sorting for agreement vs disagreement) adheres to the strict experimental method. Participants are randomly assigned between these conditions, and hence the normative views (and other characteristics) of participants are also randomly distributed across conditions. As is common in studies on normative disagreement in public good games, 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 (as per the elicitation of these participants' normative expectations); 56%

³ 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.

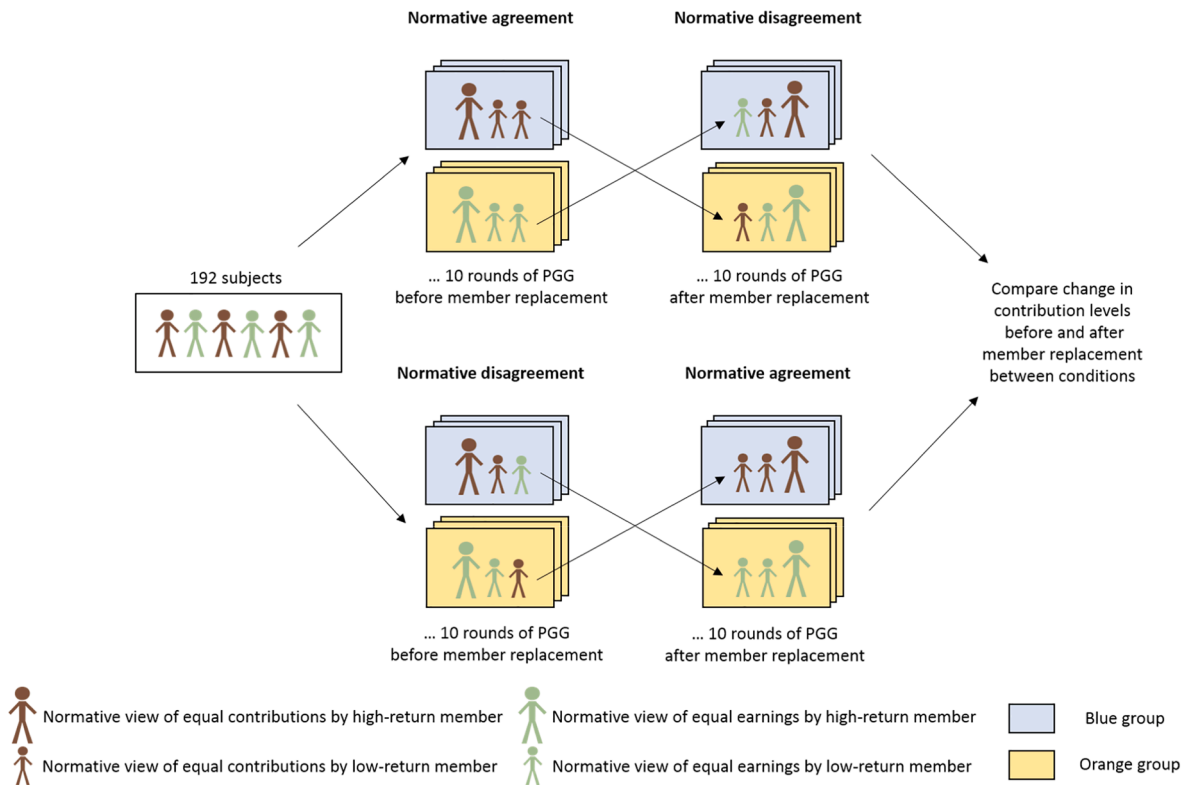
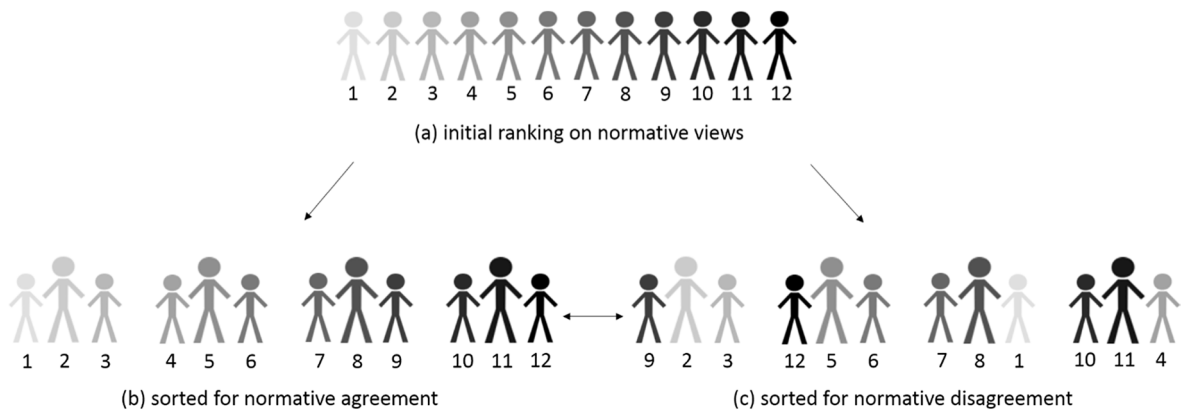


Fig. 1. Experimental design.



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.

Fig. 2. Example for method of sorting and resorting participants.

guess their group members' normative views correctly on average before the experiment, and 71% guess correctly on average after the experiment.

3.4. Social norms

Many definitions of social norms share the view that social norms involve shared expectations between group members on what actions are considered appropriate (Fehr & Gächter, 2000; Ostrom, 2000). Accordingly, mutually consistent normative expectations in a group are commonly used as an indicator of the existence of a social norm (Bicchieri, 2006; Bicchieri, Lindemans, & Jiang, 2014; Krupka & Weber, 2013). Therefore, to assess the presence 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 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 normative views, normative expectations are elicited three times: once before the first round, once after the first 10 rounds (before the membership change), and once after the last 10 rounds. 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 = 0.72, $p < .001$), we do not analyze them separately. We focus on the similarity in normative views and expectations after the last 10 rounds, but we give the similarity at all measurement moments in the supplementary material (Fig. S2).

3.5. 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

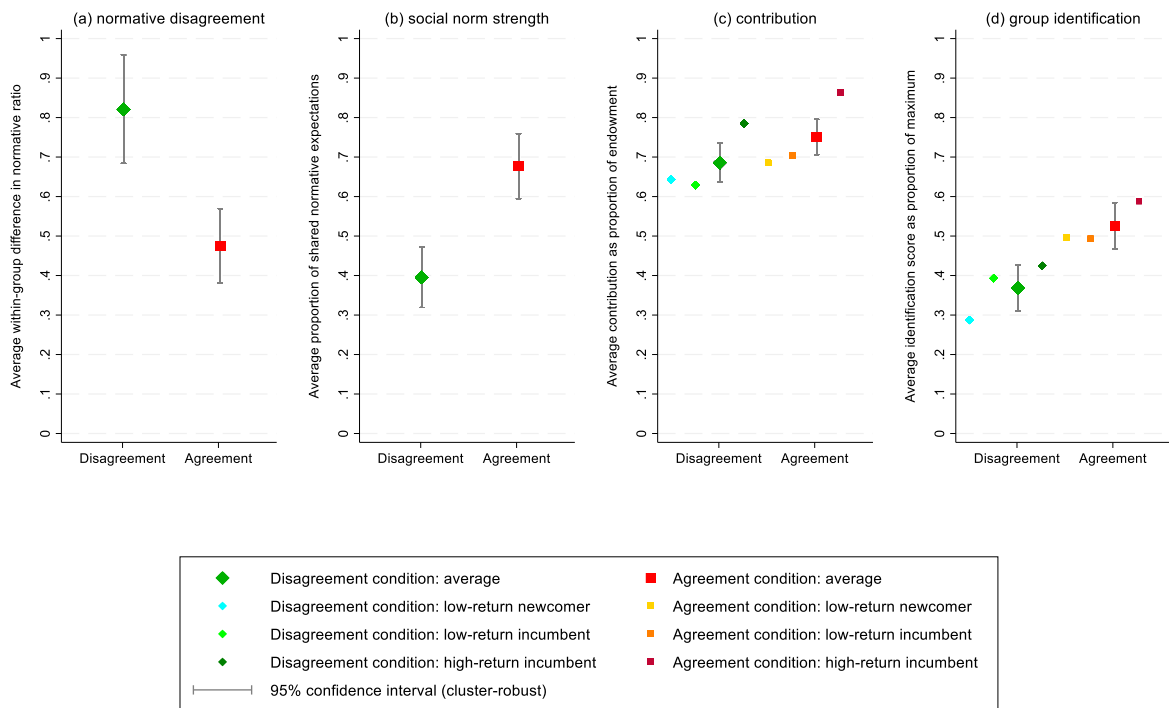
These items are commonly used to measure group identification in experiments (Leach et al., 2008; Ouwerkerk, Ellemers, & Kortekaas, 1999). We take the average across the six items as the group identification score. This score has high reliability, as indicated by a Cronbach's alpha of 0.93. We focus on the group identification score after the last 10 rounds, but we give the group identification score after the first 10 rounds in the supplementary material (Fig. S3).

3.6. Post-experiment measures and other information

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. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012). All data and code are available at <https://doi.org/10.24416/UU01-J5M2LE>.

4. Results

We first examine whether the level of normative disagreement differs by experimental treatment 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. Fig. 3a shows that groups in the condition where newcomers and incumbents are sorted for disagreement indeed end up with considerably more disagreement than groups in the condition where newcomers and incumbents are sorted for agreement (Mann-Whitney ranksum test, $p < .001$). This can also be seen in the supplementary material, Fig. S1 and S2, where we show the normative views and disagreement for each group separately and how the disagreement developed over time throughout the experiment. Fig. 3b shows that groups in the disagreement condition also end up with weaker social norms, as indicated by more dissimilarity in normative expectations among group members (Mann-Whitney ranksum test, $p < .001$). Normative disagreement between newcomers and incumbents thus also impedes social norm emergence.



Note: Disagreement refers to the condition where newcomers and incumbents are in normative disagreement. Agreement refers to the condition where newcomers and incumbents are in normative agreement. Normative disagreement (a), social norm strength (b) and group identification (d) values are based on the measurement after round 20. Contributions (c) were measured every round and we take the average of rounds 11-20.

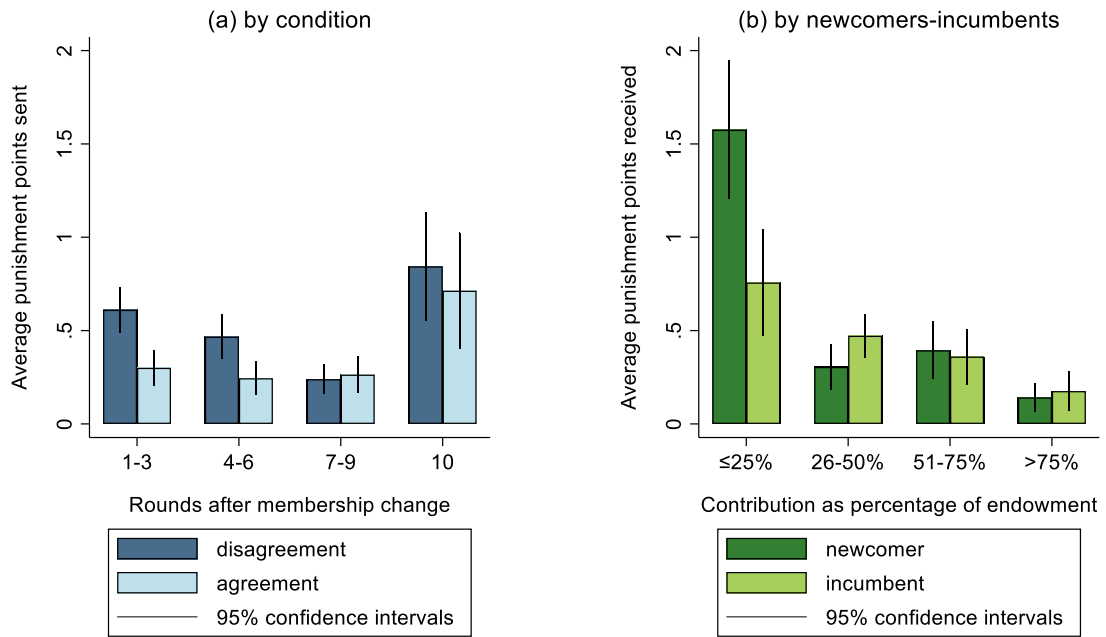
Fig. 3. Disagreement, social norms, contributions, and group identification by condition.

However, Fig. 3c shows that the extent of normative disagreement between newcomers and incumbents does not harm cooperation: the average contributions to the common good are very similar in both conditions. We find no significant differences in the contribution levels between the disagreement and the agreement conditions, neither before membership change (proportion of endowment contributed of 0.68 vs 0.71, OLS regression with cluster-robust standard errors, $p = .37$) nor after membership change (proportion of endowment contributed of 0.69 vs 0.75, OLS regression with cluster-robust standard errors, $p = .08$). Also when using population-averaged regression models, which account for repeated measures obtained from the same participant or group, we find that the change in contribution levels before and after membership change does not differ significantly between conditions, see supplementary material Table S1. The contribution levels per round can be found in supplementary material Fig. S4 and also indicate no difference between conditions and after membership change.

We additionally test the effect of normative disagreement between newcomers and incumbents conditional on contribution levels before membership change. The conditional test suggests that the change in normative disagreement brought about by the membership change does impact groups that had low contribution levels before membership change. Previously low-contributing groups do not change their contribution level after membership change in the disagreement condition, whereas previously low-contributing groups do increase their contribution level in the agreement condition (supplementary material Table S2). However, for groups that had high contribution levels before membership change, the change in contribution level after membership change did not depend on the experimental condition. Finally, we run regressions with group-level contributions as the outcome variable and the before-mentioned group-level normative disagreement as the predictor (while controlling for other factors known to influence contributions). These models indicate no significant relationship between disagreement and cooperation, also not in conditional tests (supplementary material Table S2-S3). Altogether, we find no support for the hypothesis that normative disagreements between incumbents and newcomers harm cooperation in terms of contributions to the common good.

Fig. 3d shows that normative disagreement between newcomers and incumbents does harm group identification: group identification is about 40% higher in the condition where newcomers and incumbents agree instead of disagree (Mann-Whitney ranksum test, $p < .001$). 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, $p = .04$). Regression models with group identification as the outcome variable and group-level normative disagreement as the predictor (while controlling for other factors related to group identification) also indicate a significant negative association between normative disagreement and group identification (see supplementary material, Table S4).

We next examine if normative disagreement between incumbents and newcomers affects the amount of punishment received, and whether punishment differs between incumbents and newcomers. Fig. 4a indicates that punishment levels are considerably higher in



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). Agreement refers to the condition where newcomers and incumbents are in normative agreement. Disagreement refers to the condition where newcomers and incumbents are in normative disagreement. 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 in panel (a): 192 participants * 10 rounds * 2 co-members to punish or receive punishment from. To keep the return rate constant when comparing newcomers and incumbents in panel (b), we only compare (low-return) newcomers to low-return incumbents (which make up two-thirds of the participants), giving 128 participants * 10 rounds * 2 co-members = 2560 observations. When subdividing by contribution level, the number of observations is 326 for contributions ≤ 25%, 792 for contributions 26-50%, 516 for contributions 51-75%, and 926 for contributions >75%.

Fig. 4. Punishment.

the early interactions between newcomers and incumbents in the condition with normative 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, Mann-Whitney ranksum test, $p < .001$, full details in Table S5). However, the punishment level decreases over time in the condition with disagreement and thereby also the difference between the conditions. Regressions with group-level punishment as the outcome variable and group-level disagreement as the predictor (while controlling for other factors related to punishment) do not suggest a significant association between these two variables, also not in the first rounds after membership change (see Table S6). The finding that normative disagreement between newcomers and incumbents leads to more punishment in the early interactions is thus not robust to alternative analysis. There is an outlier in terms of punishment in the last round (the tenth round after membership change), 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.

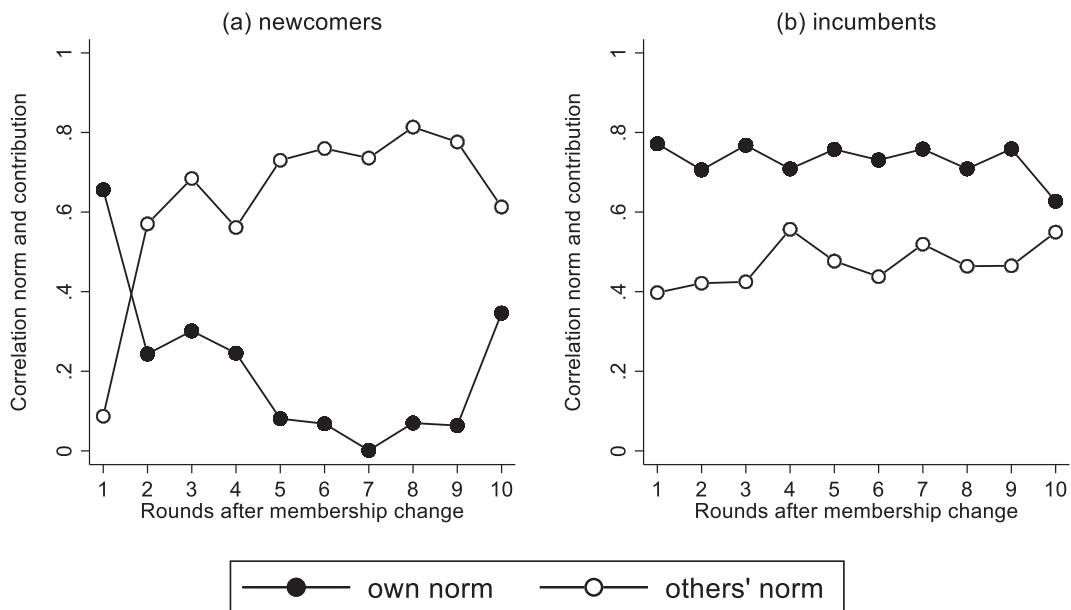
In Fig. 4b, we show the average punishment points received by incumbents and newcomers for different levels of contributions. We see that there are no significant differences between incumbents and newcomers when they contribute medium to large amounts. In this case, punishment is low for both newcomers and incumbents. However, we find that newcomers are more strongly punished for low contributions. That is, newcomers receive about double the number of punishment points when they contribute ≤25% of their endowment compared to incumbents making these contributions (1.58 vs 0.76 punishment points received, Mann-Whitney ranksum test, $p < .001$, full details in Table S5). 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 (Mann-Whitney ranksum test, $p < .001$, Table S5), but not when they are in normative agreement (Mann-Whitney ranksum test, $p = .17$, Table S5). 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 Fig. S6.

Predicting contribution levels. To examine why contribution levels do not change after membership change and what alternatively predicts contribution levels, we conduct further exploratory analyses. These full analyses can be found in supplementary material S6. Here we explain the main findings. We find that the change in the group-average normative view that is brought about by the membership change is an important factor in predicting whether the membership change increases or decreases the group-average contribution (i.e., whether the group-average contribution in the second 10 rounds is higher or lower than in the first 10 rounds).

A participant's average normative view measures what a participant deems to be the appropriate average contribution in the group. For example, if a participant thinks the high-return member should contribute 20 and the two low-return members should contribute 10, the participant's average normative view would be $(20 + 10 + 10)/3 = 13.33$. The group-average normative view is then the average of its three participants' average normative views. We find that if newcomer entry leads to a higher group-average normative view, then the group-average contribution increases, and the opposite holds when newcomer entry reduces the group-average normative view. This is not just a function of the normative view that the newcomer brings in, it also depends on the normative view of the incumbent that the newcomer replaces. The group-average normative view increases if the newcomer's view is higher than the view of the incumbent that is replaced. The change in the group-average normative view correlates at 0.53 with the change in the group-average contribution (OLS regression on 64 groups, $p < .001$, also when controlling for other related factors, Table S3). The group-average normative view is an important reason for the null-effect of membership change on contributions. When averaging over all groups (also within conditions), newcomers do not change the group-average normative view; in some groups they increase the group-average whereas in other groups they decrease the group-average, which balances out when averaging over all groups.

When predicting how contribution levels develop after membership change on an individual level, we find that a participant's own normative view and the contribution of the participant'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. Compared to incumbents, newcomers contribute more in line with the contributions of their group members and less in line with their own 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 were not punished in the prior round. But if they were punished, they increase their contribution. Moreover, the stronger the punishment, the more participants move away from their prior contribution behavior in favor of higher contributions. With high levels of received punishment (about 10 punishment points), a participant's prior contribution is no longer related to the participant's subsequent contribution. However, if participants are punished when contributing large amounts (about >15 MU), they do not significantly change their behavior (see Table S7 in the supplementary material).

Because newcomers act more in line with the contribution of incumbents, and the contribution of incumbents is related to the incumbents' normative views, newcomers end up conforming more to the incumbents' normative views than their own. The relative influence of the newcomer's own normative views and that of the incumbents is best shown in the condition with disagreement between newcomers and incumbents, because that is where they have to decide what view to conform to. In Fig. 5, we show for newcomers (a) and incumbents (b) how their contributions in each round correlate with their own normative view and the normative view of their group members. Fig. 5a shows that while newcomers start by contributing according to their own normative view in the first round after membership change, already in the second round after membership change they contribute according to their group



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 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) and correlate it with the subsequent contribution decisions.

Fig. 5. Correlation between contribution and own or others' normative views.

members' normative view. In contrast, Fig. 5b shows that incumbents contribute more in line with their own normative view throughout all rounds. These findings suggest that when there is disagreement between incumbents and newcomers on how to contribute to the public good, newcomers largely concede to incumbents. That this occurs despite our finding that normative disagreement harms social norm emergence (Fig. 3b) suggests that the contribution norms in these groups are descriptive rather than injunctive. Further analyses indicate that the finding of newcomers conceding to the incumbents cannot be attributed to newcomers changing their normative view or being a minority (see supplementary material, Fig. S7-S8).

5. 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, 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 that norm enforcement plays an important role in sustaining cooperation in groups consisting of newcomers and incumbents. Norm enforcement was possible in our experiment via punishment. Low contributing participants respond strongly to punishment by increasing their subsequent contributions, and newcomers in particular are punished strongly for making low contributions that deviate from the incumbents' contribution norm. The result is that newcomers mostly adapt to the incumbents' norm. That this occurs despite newcomers sticking to their normative views and expectations suggests that the contribution norms are descriptive rather than injunctive (Cialdini, Reno, & Kallgren, 1990).

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 realized. This also speaks to the ongoing debate on whether the integration of migrants is a one-sided or two-sided process (Klarenbeek, 2019), with our experiment providing support for the former. Similarly, the punishment results indicate a newcomer-incumbent inequality in terms of how strongly low contributors are punished. Altogether, 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. However, we did find some evidence that norm enforcement via punishment is initially higher under normative disagreement. Since punishment is costly, this does suggest some temporary harmful economic impacts of normative disagreement. This norm enforcement is especially directed at newcomers, who are punished twice as much as incumbents for contributing low amounts and hence incur extra costs. 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). These authors suggest that social control therefore will increase as more newcomers enter and communities become increasingly ethnically diverse. The norm examined in this field experiment is different from our study, namely the anti-littering norm, and the setting is very different as well (field vs lab). That we nevertheless find the same pattern of newcomers being punished more strongly than incumbents for norm violations suggests that this finding may generalize to other norms and settings.

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 public goods games 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 group relations when other norms are involved. For example, group relations between native-majority members and immigrants are often argued to be harmed by disagreements on religion or work-related norms (Hainmueller & Hopkins, 2014). Our finding that disagreement about contribution norms lowers group

identification might therefore extend to these other forms of normative disagreement, which future research can test. Moreover, we studied groups of three in which one incumbent is replaced by one newcomer. Normative disagreements might be more harmful when the number and share of newcomers are larger, as this increases the chance for separate ingroup-outgroup subdivision and the newcomers' power to influence norms (Pettigrew, 1991). Prior research suggests that if more than one newcomer enters a group, the newcomers tend to identify with each other instead of with the incumbents (Moreland, 1985), which may further lower a common sense of group identification among incumbents and newcomers. 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, Malaeb, Vlassopoulos, & Wahba, 2019). Although natural identities are conflated with potential confounders (e.g., differences in status and income), 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 perceptions (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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is part of the research program Sustainable Cooperation – Roadmaps to Resilient Societies (SCOOP). The authors are grateful to the Netherlands Organization for Scientific Research (NWO) and the Dutch Ministry of Education, Culture and Science (OCW) for generously funding this research in the context of its 2017 Gravitation Program (grant number 024.003.025).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.joep.2021.102448>.

References

- Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2021). Predicting social tipping and norm change in controlled experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 118(16). <https://doi.org/10.1073/pnas.2014893118>
- 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., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, 140(6), 1556–1581. <https://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., 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, MA: 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>
- 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>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Collier, P. (2013). The social consequences. In *Exodus: How migration is changing our world* (pp. 57–110). Oxford: Oxford University Press.
- Cubitt, R. P., Drouvelis, M., Gächter, S., & Kabilin, R. (2011). Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, 95(3–4), 253–264. <https://doi.org/10.1016/j.jpubeco.2010.10.011>
- 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>

- 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., & Laffy, 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., & Laffy, J. (2021). Social conformity under evolving private preferences. *Games and Economic Behavior*, 128, 104–124. <https://doi.org/10.1016/j.geb.2021.04.005>
- 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 (No. 6504). CESifo Working Paper.
- Fehr, E., & Gächter, 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, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7), 458–468. <https://doi.org/10.1038/s41562-018-0385-5>
- 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). *Coehtnicity: 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(1), 225–249. <https://doi.org/10.1146/annurev-polisci-102512-194818>
- Hauge, K. E. (2015). Moral opinions are conditional on the behavior of others. *Review of Social Economy*, 73(2), 154–175. <https://doi.org/10.1080/00346764.2015.1035910>
- 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>
- 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>
- Klarenbeek, L. M. (2019). Reconceptualising 'integration as a two-way process'. *Migration Studies*, 1–20. <https://doi.org/10.1093/migration/mnz033>
- 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., ... 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>
- 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), e55058. <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](https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/3<371::AID-EJSP932>3.0.CO;2-U)
- 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 Series No. 4719). Retrieved from <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*. Berlin, Boston (pp. 239–256).
- Reuben, E., Riedl, A., & Bernard, M. (2015). Fairness and coordination: The role of fairness principles in coordination failure and success. Retrieved from <https://sites.google.com/site/markbernard1984/research-papers>.
- 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>
- 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>
- Salmon, T. C., & Weber, R. A. (2017). Maintaining efficiency while integrating entrants from lower performing groups: An experimental study. *Economic Journal*, 127(600), 417–444. <https://doi.org/10.1111/eoj.12308>
- 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: London: Palgrave Macmillan. <https://doi.org/10.1057/9780230248359>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology* 26, pp. 4–7. DOI: 10.2139/ssrn.2160588.
- Smerdon, D., Offerman, T., & Gneezy, U. (2019). 'Everybody's doing it': on the persistence of bad social norms. *Experimental Economics*. <https://doi.org/10.1007/s10683-019-09616-z> (0123456789).
- 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](https://doi.org/10.1016/s0165-1765(98)00203-1)
- Spadaro, G., Tiddi, I., Columbus, S., Jin, S., Teije, A. ten, & Ballieta, D. (2020). The Cooperation Databank.

- 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.
- 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: Oxford: Basil Blackwell. DOI: 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>