Methods

Experiment

We conducted the computerized experiment in the Experimental Laboratory for Sociology and Economics (ELSE) at Utrecht University during October-November 2019. The experiment was programmed with z-Tree software (Fischbacher, 2007). We recruited participants amongst students at Utrecht University using the internet recruitment system ORSEE (Greiner, 2015). We ran 8 sessions with 24 participants each, leading to a total of 192 participants. Each session lasted about 75 minutes. Payment depended on behavior in the game, participants earned on average 15 euros (min = 5, max = 22). Participants were on average 24 years old, 127 (66%) were female, 62 male, and 3 other. Almost all participants were students at Utrecht University, 87 were Dutch and 105 from various other countries. The participants play 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 use to adapt it for our purposes.

The Game

Each round of the public good game with peer punishment has two stages. In the first stage, each individual i in a group composed of N members receives an endowment E, and must decide how much of this endowment to contribute to a public good, c_i, where $c_i \in \{0, 1, \dots, E\}$. The part of the endowment that is not contributed to the public good is kept for the individual. The public good consists of the sum of the contributions made by all individuals. Each individual receives a return per contributed point (sometimes also referred to as marginal per capita return) to the public good ($m_i < 1$). The sum of these returns makes up the total multiplication factor of the public good M, with $N > M > 1$. Because the individual return of a contribution is always smaller than 1, it is most profitable for the individual to contribute nothing. Because the total group return (multiplication factor) is always bigger than 1, it is most profitable for the group (i.e., everybody combined) if everybody contributes fully. These two aspects together form the social dilemma of public good provision. After all individuals in a group have made their contribution decision, the contributions and payoffs of each player are communicated to all and the first stage is finished.

In the second stage, each individual is given the opportunity to assign punishment points $p_{ij} \in \{0, 1, \dots, \max(p_{ij})\}$ to each group member j. Each punishment point costs 1 point, to the

punisher, and reduces the payoff of the punished player by δ points. The individual payoff after one round of this two-stage is given by:

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

Individuals do not see who punished them (to prevent confounding of normative behavior with revenge motives), and groups repeatedly play rounds of this two-stage game within the same group.

We induce the possibility of opposing normative views by assigning heterogeneous returns of the public good to the participants. Specifically, we assign per group one participant with a high return m_i = .75, and two participants with a lower return m_i = .50 (making the total multiplication factor M = 1.75). Prior studies suggested that with the same heterogeneity in returns there is considerably disagreement in personal normative views (e.g., Reuben & Riedl, 2013); the view that everybody should contribute fully is endorsed by some, the view that high-return actors should contribute more than low-return actors in such an extent that final earnings are equal is endorsed by others, and still others deem a mix between the two appropriate (i.e., high-return actors should contribute more than low-return actors, but not to such an extent that earnings are equalized).

For comparability to previous research, the rest of the parameter values are set to follow the typical form of the public good game with peer punishment (Fehr & Gachter, 2000; the endowment E = 20, the impact of receiving punishment δ = 3, and maximum punishment per group member p_{ijmax} = 10). Table 1 indicates the predicted contributions and resulting payoffs for each type of participant that match the full contribution rule and the equal earnings (relative contribution) rule.

Table 1. Contributions and Earnings under Full Contribution and Equal Earnings Rule

	c_H	c_L	π_H	π_L
Full contribution	20	20	45	30
Equal earnings	20	10	30	30

c_H = contribution of participant with high return, c_L = contribution of participant with low return, π_H = payoff for participant with high return, π_L = payoff for participant with low return.

Procedure

Prior to play, we present the participants with the choice situations, and then measure for each participant the personal normative view on what he/she deems an appropriate contribution rule (as will be explained later). We implement two conditions, that differ only in the method of (re)sorting participants in groups based on their normative views.

In the first condition, we sort participants of similar normative views together, while in the second condition, we sort participant of dissimilar normative views together. In both conditions, the participants play 20 rounds in total of the public good game. The first 10 rounds they play within their group. After the first 10 rounds, we exchange one member per group for a member from another group in both conditions, such that each group receives a newcomer in place of an old-timer, and let the newly formed groups play the second set of 10 rounds. The resorting is done in such a way that the (dis)similarity in normative views reverses between conditions; the groups in the condition with initial sorting on similar normative views becomes as dissimilar (in terms of the first elicited normative views) as the groups in the condition with initial sorting on dissimilar normative views, and vice versa. We examine whether the entry of newcomers who are accustomed to a different social norm than the old-timers impede cooperation towards public good provision, and how this differs from the situation where newcomer entry brings in a normative view that matches those of the old-timers.

At three moments in the game we elicit the personal normative view of each participant. The first is prior to the first 10 rounds of the game, as we already mentioned. The second is prior to the second 10 rounds of the game, and the third is after the second 10 rounds of the game. In this way, we try to assess the stability of personal normative views. To examine the development of social norms, we also elicit the normative expectations of each participant at these three moments. We now turn to the measures for these attributes.

Elicitation of Normative Views

We directly elicit the participant's personal normative view on what he/she deems an appropriate contribution rule to the public good. Prior to sorting participants into groups and assigning them their returns, we present them with the game they will play and elicit their normative view.[1] We do this by letting the participant indicate what he/she thinks to be the

[1] We opt for elicitation before the subjects know their return from the public good, because that is their real unbiased view of how to contribute. Moreover, sorting on similar normative views that have been elicited after subjects know their return could be problematic if there is a large self-serving bias, i.e., high-return types almost always prefer equal contributions while low-return types almost always prefer equal earnings.

appropriate contribution decisions for another hypothetical group of participants. The participant is asked to indicate the appropriate contribution of each of three group members, one with return m_i = .75, and two with m_i = .50. The participant can try out different combinations of contributions, and see how it affects the earnings of each group member. The participant's decision is anonymous.

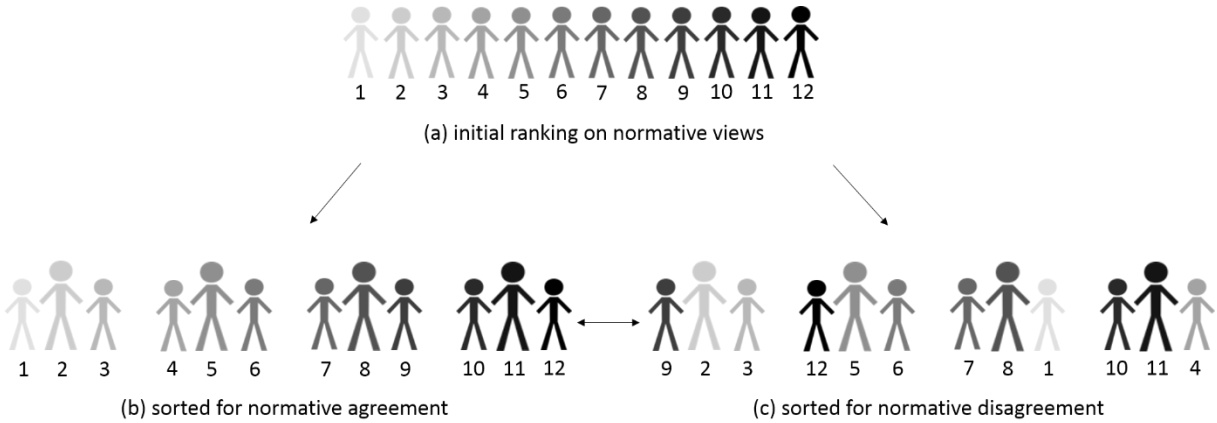
Sorting on Normative Views

The direct elicitation of the participant's normative view not only allows us to examine how behavior in the game is associated with the normative view, but also to manipulate at the group-level whether there is similarity or not. As mentioned, we compare two conditions that differ in the method of sorting participants in groups. In one condition, we sort participants into groups that are similar in terms of their normative views. In the other condition, we sort participants into groups that are dissimilar in terms of their normative views. To do so, we assign the participant a ranking in terms of how much of how much he/she cares about the equal earnings rule versus the equal contributions rule compared to the other participants. With an equal earnings rule, the contribution of the member with a high return c_H should be larger than the contributions of members with a lower return c_L, while they should be equal with an equal contribution rule. The extent to which the normative view is catered towards equal earnings is therefore captured by the score: $c_H - \overline{c_L}$, and participants are ranked according to this score.[2,3] The method of sorting and resorting based on these ranks is described in Figure 1.

[2]More precisely, the ranking score is: $c_H - \overline{c_L} + 0.02\,\overline{c} + 0.0001R$, where R is a random number between 0 and 1. The addition of $0.02\,\overline{c}$ makes sure that persons that assign a contribution of 20 to all members obtain slightly higher scores than persons that assign a contribution of 0 to all members. This helps to differentiate between those two possible ways of assigning contributions in the sorting method in case multiple persons within a session have such norms. 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 mean level of contributions. The addition of $0.0001R$ is to avoid tied scores.

[3] Although unlikely, it is possible that people will assign higher contributions to actors with a lower return, which would indicate a normative view that matches neither equal contributions nor contributions relative to returns. In the condition with sorting on similar normative views, we group subjects with such normative views together if there are multiple of them, and otherwise we group them together with the subjects catered more towards equal contributions. In the analyses, we will assess whether the results differ if groups with such individuals are left out.

Figure 1. Example for Method of Sorting and Resorting Participants



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 it 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 stick figures.

We do not tell participants about the method of sorting. Instead, participants are informed when the sorting happens (directly after the norm elicitation), but now how it happens.

Elicitation of normative expectations

What the participant personally believes is appropriate to do (normative views), and what the participant believes that others deem appropriate to do (normative expectations, or perceived social norms) may not be the same, and can both have an influence on behavior that is to some extent independent of the other. We obtain a measure for normative expectations in the following way.

After the participant reported his/her personal normative view, we tell him/her that his/her group members were also asked to indicate appropriate contributions for three members in the public good game. The participant is then asked to guess the answers submitted by his/her group members. To incentivize the guess, the participant is told that we randomly pick one of his/her guesses, and give an additional payment of 100 MU when it matches the answer of at least one of the group members. Only at the end of the experiment are participants informed of whether they were correct in the guess we randomly chose. This measure is an adaptation of

(Krupka & Weber, 2013) similar to that of (Reuben et al., 2015). Mutually consistent normative expectations in a group are a sign that a social norm exists (Bicchieri et al., 2014).

3.6 Punishment norms

Because we allow for peer punishment, norms of punishment may also influence cooperation, and possibly moderate the influence of normative (dis)agreement on contributions. To assess this possibility, we ask participants after the 20 rounds of the game to give their own view, and their expectation of the other participants' view, on the appropriate number of punishment assigned to each member in a hypothetical group of three in two contexts. In the first context, one low-return participant contributes 20 points, while the other two participants (one low- and one high-return) contribute 0. In the second context, two low-return participants contribute 10 points, and one high-return participant contributes 20 points. The expectations on the normative views are incentivized in the same way as described for normative expectations for contributions.

3.7 Group identification

Both after the first and second set of 10 rounds, we ask each participant to indicate his/her 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

3.8 Conditionality of normative views

At the end of the experiment, we measure how conditional the participant's normative views are. We ask the participant to indicate the appropriate contribution for a low-return group member when the other two group members (one low-return with $m_i = .50$ and one high-return with $m_i = .75$) contributed on average 0 points, 2 points, 4 points, and so on until 20 points. We also ask the participant to indicate the appropriate contribution for a high-return member under

these average contributions of the other two group members (both members with a low return of m_i = .50).

3.9 Social preferences and background questions

After the measure on the conditionality of normative views, we elicit social preferences to see to what extent they associate with normative views and other background questions. To measure social preferences, we use the z-tree implementation (Crosetto et al., 2012) of the Social Value Orientation Slider Measure by (Murphy et al., 2011). The measure is a composite of six primary and nine secondary items, and we include all items. Each item requires the participant to allocate a payoff between oneself and another participant. The payoffs between the 15 items are varied such that the motives of altruism, individualism, competition, efficiency, and inequity aversion can be distinguished.

The background questions measure age (open), gender (male, female, other), nationality (open), field of study (open, only asked if participant is a student), and political orientation (10 point scale from left-wing to right-wing). We also ask the participants to rate their understanding of the experimental instructions (bad / not bad, not good / good), in how many experiments they participated at the lab, how many participants in the room they know by first name, and if they followed a course in game theory. Finally, we give participants in open-text boxes the option to communicate what they think was the purpose of the experiment, whether they associate the contribution decisions of the experiment with any real life situations, and any other comments they may have.

References

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

Crosetto, P., Weisel, O., & Winter, F. (2012). A flexible z-Tree implementation of the Social Value Orientation Slider Measure (Murphy et al. 2011). In *Jena Economic Research Papers* (No. 062). https://doi.org/10.1016/s0749-0690(05)70105-1

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

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

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

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

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781. https://doi.org/10.2139/ssrn.1804189

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*. https://sites.google.com/site/markbernard1984/research-papers