White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/195962/

Version: Accepted Version

Article:
Grigoryan, Lusine, Seo, San, Simunovic, Dora et al. (2023) Helping the ingroup versus harming the outgroup: Evidence from morality-based groups. Journal of Experimental Social Psychology. 104436. ISSN: 0022-1031

https://doi.org/10.1016/j.jesp.2022.104436

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Helping the ingroup versus harming the outgroup: Evidence from morality-based groups

Lusine Grigoryan[1], San Seo[1], Dora Simunovic[2], Wilhelm Hofmann[1]

[1] Faculty of Psychology, Ruhr University Bochum

[2] Bremen International Graduate School of Social Sciences (BIGSSS), Jacobs University Bremen

Author note

Lusine Grigoryan https://orcid.org/0000-0002-2077-1975

Dora Simunovic https://orcid.org/0000-0001-8491-7227

Wilhelm Hofmann https://orcid.org/0000-0003-0295-4679

Correspondence should be addressed to Lusine Grigoryan, University of York, UK. Email: lusine.grigoryan@york.ac.uk.

Acknowledgments: This work was supported by the EASP 2021 Seedcorn Grant and the DFG individual grant #464524346.

We would like to thank Bethan Hannah Jones for her help in data collection for Study 2 and Vladimir Ponizovskiy for helpful discussions and feedback.

This paper has been accepted for publication at *Journal of Experimental Social Psychology* in December 2022. This paper is not the copy of record and may not exactly replicate the authoritative document published in JESP.

Citation: Grigoryan L., Seo S., Simunovic D., Hofmann W. (2023). Helping the ingroup versus harming the outgroup: Evidence from morality-based groups. *Journal of Experimental Social Psychology*, 105. https://doi.org/10.1016/j.jesp.2022.104436

Abstract

The discrepancy between ingroup favoritism and outgroup hostility is well established in social psychology. Under which conditions does "ingroup love" turn into "outgroup hate"? Studies with natural groups suggest that when group membership is based on (dis)similarity of moral beliefs, people are willing to not only help the ingroup, but also harm the outgroup. The key limitation of these studies is that the use of natural groups confounds the effects of shared morality with the history of intergroup relations. We tested the effect of morality-based group membership on intergroup behavior using artificial groups that help disentangling these effects. We used the recently developed Intergroup Parochial and Universal Cooperation (IPUC) game which differentiates between behavioral options of weak parochialism (helping the ingroup), strong parochialism (harming the outgroup), universal cooperation (helping both groups), and egoism (profiting individually). In three preregistered experiments, we find that morality-based groups exhibit less egoism and more universal cooperation than non-morality-based groups. We also find some evidence of stronger ingroup favoritism in morality-based groups, but no evidence of stronger outgroup hostility. Stronger ingroup favoritism in morality-based groups is driven by expectations from the ingroup, but not the outgroup. These findings contradict earlier evidence from natural groups and suggest that (dis)similarity of moral beliefs is not sufficient to cross the boundary between "ingroup love" and "outgroup hate".

Keywords: cooperation, ingroup bias, prejudice, morality, economic game

Helping the ingroup versus harming the outgroup: Evidence from morality-based groups

"Ingroup love" does not equal "outgroup hate". People are quick to favor the ingroup, but reluctant to harm the outgroup – a phenomenon well established in experimental studies with both artificial and natural groups (Balliet et al., 2014; Brewer, 1999; Buhl, 1999; Hewstone et al., 2002; Mummendey & Otten, 1998). Under which conditions does ingroup love turn into outgroup hate? Group members' belief in the ingroup's moral superiority has been proposed as one such condition (Brewer, 1999; Ellemers & van den Bos, 2012). Experiments with natural groups show that morality-based groups exhibit more outgroup hostility than non-morality-based groups (Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). Natural groups, however, come with a long history of intergroup relations, including polarization and conflict, making it nearly impossible to disentangle the role of this history from the effect of shared morality itself. The aim of the current research is to test whether shared morality is indeed a sufficient condition for the emergence of outgroup hostility in intergroup cooperation settings using artificial groups with no prior history of group identification or intergroup conflict.

Shared morality is one of the core aspects of social identity (Ellemers & Van der Toorn, 2015). However, social groups differ in the extent to which morality is central to the group membership (Parker & Janoff-Bulman, 2013). We define morality-based groups as groups that are differentiated by (dis)similarity of their moral beliefs. Religious and political groups would be typical examples of morality-based groups, but any group can be perceived as a morality-based groups if the group is considered to be informative of its members beliefs and values. Perceptions of moral dissimilarity are responsible for much of hostility and polarization that exists in many societies today (Arvan, 2019; Brandt & Crawford, 2019; Kovacheff et al., 2018). To effectively communicate and cooperate across political, national, and religious boundaries, we need to understand which specific properties of morality-based

groups elicit negativity towards outgroups. Here, we make the first step towards addressing this broader question, by testing whether the mere knowledge that group membership is defined by (dis)similarity of moral beliefs is sufficient to elicit hostility towards the outgroup.

Cooperation and Discrimination in Morality-Based Groups

From an evolutionary perspective, cooperative behavior in humans would be selected and maintained only if it directly or indirectly benefits the individual or their group (Buss, 1987; West et al., 2011). For cooperation to be beneficial for humans, there need to be mechanisms in place to ensure that one is not cheated in social exchange. Ingroup bias and stigmatization function as such mechanisms of control (Kurzban & Leary, 2001). Social groups with clear boundaries and normative prescriptions that guide group members' behavior ensure the maintenance of intragroup cooperation (Brewer & Caporael, 2006). Accordingly, people expect ingroup members to cooperate more (Romano et al., 2017) and cooperation with ingroup members seems more predictable and less risky than cooperation with outgroups (Brewer & Caporael, 2006). Meta-analytical evidence suggests that individuals tend to cooperate more with ingroup than with outgroup members, and this effect is consistent across natural and experimentally manipulated groups (Balliet et al., 2014).

The preference for cooperation with ingroup members arises from perceptions of predictability and trust. Groups based on commonality of moral beliefs and norms are particularly efficient in fostering trust, and therefore should give rise to stronger ingroup bias than groups based on other criteria. Evidence in support of this intuition is ample. Moral traits dominate impression formation both in interpersonal and intergroup settings (Brambilla et al., 2011, 2012; Goodwin et al., 2014). Perceptions of morality are a more important predictor of positive ingroup evaluations than perceptions of sociability and status (Leach et al., 2007). Moral diversity is less desirable than demographic diversity, and moral differences increase desired social distance more than demographic differences (Haidt et al., 2003). When information about others' multiple group memberships is available, participants show stronger

preference for ingroup members on morality-based, belief-indicative dimensions of social categorization (e.g., religion) compared to neutral (e.g., age) or status-indicative dimensions (e.g., education) (Grigoryan, 2020b, 2020a; Grigoryan, Cohrs, et al., 2022; Grigoryan, Jones, et al., 2022).

There is growing evidence that morality-based groups do not only show stronger preference for the ingroup, but also more hostility to outgroups. In interpersonal settings, attitude dissimilarity can lead to repulsion (Rosenbaum, 1986) and moral disagreement evokes anger (Mullen & Skitka, 2006; Skitka & Mullen, 2002). When interpersonal conflict is framed as a conflict of values versus conflict of interests, participants show more self-involvement and perceive less common ground, which can lead to further conflict escalation (Kouzakova et al., 2012). In intergroup settings, evidence from cross-sectional studies shows that worldview conflict consistently predicts stronger prejudice across target groups (Brandt & Crawford, 2019).

Only few experimental studies so far directly tested the effect of shared morality on intergroup attitudes and behavior. Parker and Janoff-Bulman (2013) used attitudes to abortion (Studies 1 and 2) and political affiliation (Study 3) as criteria for morality-based group membership. Compared to non-morality-based groups (sports clubs and university affiliation), morality-based groups showed more negative emotions towards the outgroup. Weisel and Böhm (2015) quasi-experimentally manipulated the degree of enmity between the groups and shared morality by using affiliation with political parties in Germany. Participants played different variants of the Intergroup Prisoner's Dilemma - Maximizing Differences (IPD-MD) game. The morality-based groups were more likely to harm the outgroup than both weak and strong enmity groups. In both cases, morality as basis for group membership was manipulated using real societal issues that participants are likely to feel strongly about. To test whether shared morality itself is sufficient for the emergence of outgroup hostility, we conduct a series of experiments with minimal groups that help disentangle the effects of shared morality from

the numerous confounding factors that come with the use of natural groups, such as history of intergroup polarization and conflict.

Current research

Until recently, behavioral measures assessing intergroup cooperation and discrimination were not able to differentiate between ingroup favoritism, outgroup hostility, and universal cooperation. The Intergroup Prisoner's Dilemma – Maximizing Differences (IPD-MD, Halevy et al., 2008) allows participants to choose between helping the ingroup or harming the outgroup, but does not provide an option for universal cooperation – benefiting both the ingroup and the outgroup. When morality concerns are salient, the desire to cooperate might transcend group boundaries. When the study design does not offer the option of universal cooperation, participants would most likely choose weak parochialism instead, artificially increasing the estimates of ingroup favoritism. The Nested Social Dilemma (NSD; Wit & Kerr, 2002) allows participants to choose between the local pool (helping the ingroup) or the global pool (universal cooperation), but does not provide an option for outgroup hostility. The newly developed Intergroup Parochial and Universal Cooperation (IPUC) game (Aaldering & Böhm, 2020) addresses this limitation. The IPUC provides a comprehensive set of behavioral options, allowing participants to either keep their endowment (egoism), to benefit the ingroup without harming the outgroup (weak parochialism), to benefit the ingroup while harming the outgroup (strong parochialism), and benefit both groups equally (universal cooperation).

Since morality can be viewed as the solution to the problem of cooperation ("morality-as-cooperation" hypothesis; Curry et al., 2019), we expect that morality-based groups will elicit less egoism, and therefore more cooperative behavior than non-morality based groups (H1). Consistent with evidence on stronger ingroup bias in morality-based groups (Grigoryan, Cohrs, et al., 2022; Haidt et al., 2003), we expect morality-based groups to elicit more weak parochialism than non-morality-based groups (H2). Consistent with emerging evidence on

negativity towards morality-based outgroups (Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015), we expect morality-based groups to also elicit more strong parochialism than non-morality-based groups (H3). Based on the well-established relationship between trust and cooperation (Balliet & Van Lange, 2013; Deutsch, 1962), we expect that the link between morality-based group membership and parochialism will be partially explained by trust (H4). We report all measures, manipulations, and exclusions in all studies.

Pilot study

The goal of the pilot study was to pretest the experimental procedure and obtain an estimate of the effect size for determining the required sample size for the main study.

Method

Participants

Power analysis was conducted in G*Power (Faul et al., 2007). To find a medium-sized global effect (f^2 = .15) in MANOVA with a power (1-β) of .80 and α = .05, the desired sample size was N = 86. We recruited more participants (N = 147) to ensure that enough valid responses will remain after removal of missing and invalid data. We did not analyze the data until the data collection was completed. Eighteen participants were excluded as they incorrectly answered more than three out of six comprehension questions for the IPUC game. The effective sample size was N = 129. Participants were predominantly students (90%) at [Anonymized] University in Germany. Seventy one percent were women, 28% men, 1% non-binary. Age varied between 18 and 53 years old, M = 23, SD = 5.9.

Students received one credit point for participation. Additionally, two €50 prizes were raffled. Participants received lottery tickets based on the outcomes of the game: the more points they obtained, the more lottery tickets they got. The winners were selected randomly, one from the control condition and one from the experimental condition.

Design and procedure

The study procedure received ethical approval from the Psychology department of [Anonymized] University. The experiment was conducted online and hosted on the survey platform Qualtrics (Qualtrics, 2005). The data was collected between November and December 2020. All participants gave informed consent. The preregistration protocol is available at https://osf.io/nu28p/?view_only=dffc166ac74344cca43e8437d520d046.

Manipulation. Participants were randomly assigned to either morality condition or the control. In the morality condition, they filled out the Oxford Utilitarianism Scale (Kahane et al., 2018) and were randomly assigned either to the group of "utilitarians" or "non-utilitarians". In the control condition, they estimated the number of Xs on the screen and were randomly assigned either to the group of "overestimators" or "underestimators". After the group assignment task, participants played the IPUC game and completed a series of questionnaires. The order of the game and the questionnaires was counterbalanced.

As a manipulation check, we developed a measure to assess how informative participants believed the group membership to be for inferring group members' beliefs and status. A successful manipulation should increase perceptions of belief-indicativeness of the group and have no effect on perceptions of status-indicativeness. We asked participants what they can tell about someone if the only thing they knew about that person was their group membership (using the label for the outgroup). The belief-indicativeness (BI) and status-indicativeness (SI) of the group membership were measured with three items each (BI: $\alpha = .71$, e.g., "Can you tell what kind of values this person has?"; SI: $\alpha = .75$, e.g., "Can you tell how successful this person is?").

IPUC. Participants were told that they will be playing in a group of four utilitarians/ non-utilitarians (experimental condition) or overestimators/ underestimators (control condition) with another group of four (the opposite label). The game was presented as a "Group decision-making task", and participants were informed that they will need to make a split decision and that their financial gain (monetary units that would later be translated into

lottery tickets) will result from the combination of their own decision and the decisions of other study participants. After receiving detailed game instructions (German version adopted from Aaldering & Böhm, 2020), participants could choose how to distribute their endowment of 10 monetary units across the four pools (A to D). The pattern of returns for contributions is presented in Table 1. Participants then answered to six comprehension questions and made their contributions once (see Aaldering & Böhm, 2020, Experiment 3). To make the contributions, participants typed the number of monetary units they wanted to contribute to each of the pools (the sum of all contributions had to equal 10). The contributions page included the summary of rules (Table 1) as a reminder.

Table 1

Pattern of Returns for Each Unit of Contribution in IPUC

	Weak parochial cooperation (Pool A)	Strong parochial cooperation (Pool B)	Universal cooperation (Pool C)	Egoism (Pool D)
Participant	+0.5	+0.5	+0.4	1
Each ingroup member	+0.5	+0.5	+0.4	0
Each outgroup member	0	-0.25	+0.4	0

Measures. The questionnaire included measures of ingroup identification, ingroup and outgroup attitudes, perceptions of threat from the outgroup, and perceived group entitativity and essentialism. These additional measures are reported in the Supplemental Online Materials (SOM, Table S1).

Results and Discussion

The manipulation was only partially successful: As expected, morality-based groups were seen as more indicative of their members' beliefs than non-morality-based groups: $M_m = 2.4$, $SD = 0.7$ vs. $M_{nm} = 1.8$, $SD = 0.8$, $d = 0.85$, $F\ (1, 126) = 23.3$, $p < .001$. Unexpectedly, groups in the control condition were perceived as more informative of their

members' status: M_m = 1.5, SD = 0.6 vs. M_{nm} = 1.9, SD = 0.9, d = -0.65, F (1, 126) = 13.6, $p < .001$.

We found no significant effect of condition on contributions in a one-way MANOVA: $f^2(V)$ = .06, η^2 = .06, *95% CI* [.00, .13], $F_{approx.}$ (3, 125) = 2.5, p = .065. After controlling for false discovery rate using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995), contributions to none of the pools differed significantly by condition. Descriptively, morality-based groups, compared to non-morality-based groups, exhibited less egoistic behavior (M_m = 1.7, SD = 2.3 vs. M_{nm} = 2.9, SD = 3.0), more weak parochialism (M_m = 2.9, SD = 2.5 vs. M_{nm} = 2.5, SD = 2.2), and more universal cooperation (M_m = 3.6, SD = 3.4 vs. M_{nm} = 2.7, SD = 3.3). Both groups showed low levels of strong parochialism: M_m = 1.8, SD = 2.7 vs. M_{nm} = 1.9, SD = 2.3.

To understand whether absence of the expected effects was due to small sample size or due to an unsuccessful manipulation, we next explored whether the contributions differed by the randomly assigned group labels (utilitarian/ non-utilitarian, overestimator/ underestimator). There were no significant differences in contributions between "utilitarians" and "non-utilitarians": $f^2(V)$ = .004, $F_{approx.}$ (3, 63) = 0.08, p = .970. However, there were significant differences in contributions between "overestimators" and "underestimators": $f^2(V)$ = .15, $F_{approx.}$ (3, 58) = 3.38, p = .024. Specifically, "underestimators" contributed more to the weak parochialism pool than "overestimators" did: M_u = 3.3, SD = 2.3 vs. M_o = 1.6, SD = 1.9, F (1, 60) = 9.9, p = .003. Since participants in this condition perceived the groups as more indicative of their members' status, it is likely that the prefixes "over" and "under" created associations with higher and lower status, and the lower-status group exhibited more weak parochialism than the higher-status group.

This pilot study identified some important methodological limitations. First, the observed effect size of the MANOVA global effect was considerably smaller than initially anticipated. Second, the manipulation had several weaknesses. The specific labels used for

group assignment created unexpected differences in perceptions of group status and the utilitarianism scale was a relatively weak manipulation for morality-based group membership. These limitations were addressed in Study 1.

Study 1: No Evidence of Strong Parochialism in Morality-Based Groups

Study 1 aimed to test the effect of morality-based vs. non-morality-based group membership on cooperative intergroup behavior using the IPUC game. We predicted that morality-based groups, compared to non-morality-based groups, would exhibit less egoism (H1) and more weak parochialism (H2). Since our pilot data provided some preliminary evidence that morality-based groups might contribute more to the universal cooperation pool rather than the strong parochialism pool, we preregistered two competing hypotheses: morality-based groups will show more strong parochialism (H3a) or more universal cooperation (H3b). The preregistration protocol is available at https://aspredicted.org/YLN_AVK.

Method

Participants

Based on the observed effect size for the overall distribution of contributions in the pilot study (*Pillai's V* = .06), the power analysis indicated a desired sample size of *N*=296 ($\alpha = .05$, $1\text{-}\beta = .95$). As before, we sampled more participants to make up for potential missing and invalid responses. Participants were recruited on Prolific Academic and among the students of the [Anonymized] University in Germany. Participants on Prolific received £1.88 for participation and students received one credit point. Two €50 prizes were raffled in both participant pools, with same instructions as in the pilot study.

A total of 456 participants completed the questionnaire. We did not analyze the data until the data collection was completed. We excluded 59 participants who failed more than three of the six comprehension questions for IPUC. One additional participant was excluded as they made contributions that were outside of the allowed 0–10-unit range. The effective

sample size was $N = 396$. Participants were predominantly women (60% women, 39% men, 1% non-binary) and 64% were students. Age varied between 18 and 69 years old, $M = 26.3$, $SD = 8.3$.

Design and Procedure

We made several changes to the experimental procedure to address the limitations identified in the pilot study. We used the Moral Foundations Questionnaire (MFQ, Graham et al., 2011) instead of the Oxford Utilitarianism Scale (OUS) for the morality-based group division, as the OUS captures only one specific aspect of moral beliefs (utilitarianism), whereas the MFQ captures a range of beliefs that are predictive of multitude of other attitudes (Graham et al., 2013). We expected the MFQ to be a stronger manipulation for morality-based group division than the OUS. In the experimental condition, participants filled out the Moral Foundations Questionnaire (Graham et al., 2011) and irrespective of their scores, were assigned to a neutrally labeled group X. They were informed that they will be playing with three other participants who use *similar criteria to judge what is right and what is wrong*, with another team of four (group Y), who use different criteria to judge what is right and what is wrong. In the control condition, participants were presented with five pairs of paintings and asked to choose the one they liked more. Irrespective of their choices, they were assigned to group X that consisted of four people who have *similar taste in art,* playing with group Y that consisted of 4 people who had a different taste in art.

The IPUC game instructions and comprehension questions were identical to the pilot study, only the names of the groups were changed. The order of the game and the measures were counterbalanced. We used the same manipulation checks as in the pilot study (belief-indicativeness: $\alpha = .91$, status-indicativeness: $\alpha = .78$). Additional measures included identification, ingroup and outgroup attitudes, perceived group entitativity and essentialism (see SOM, Table S2). Finally, we included several items to measure trust and expectations from ingroup and outgroup members. Participants were asked to indicate their agreement with

two items: "I trust members of [my Group X / the other group Y] to cooperate in this game" on a 7-point scale from *'Absolutely disagree'* to *'Absolutely agree'*. We also asked how likely they thought it is that the ingroup and outgroup members would contribute to each of the pools on a 10-point scale from *'Very unlikely'* to *'Very likely'*.

Results and Discussion

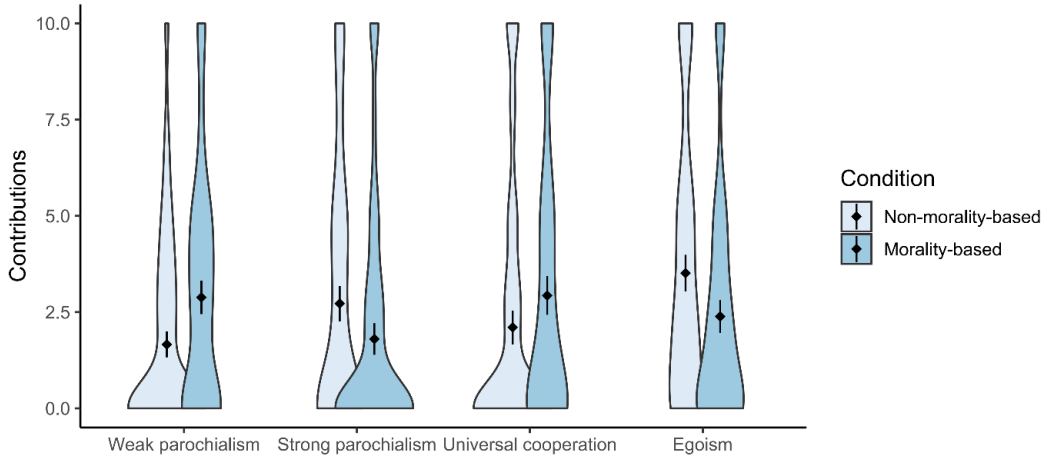
The manipulation was successful: participants in the morality-based condition perceived the groups to be more informative of their members' beliefs: M_m = 2.5, SD = 1.1 vs. M_{nm} = 1.3, SD = 0.6, d = 1.39, F (1, 393) = 190, p < .001, but not status: M_m = 1.3, SD = 0.6 vs. M_{nm} = 1.2, SD = 0.5, d = 0.17, F (1, 393) = 2.99, p = .085.

Preregistered hypotheses

There was a significant effect of condition on the overall distribution of contributions: $f^2(V)$ = .08, η^2 = .08, *95% CI* [.03, .13], $F_{approx.}$ (3, 392) = 10.9, p < .001. Fig. 1 shows the differences in contributions to each pool and table S3 in the SOM reports the mean and median contributions to each pool. All p-values were adjusted for the false discovery rate. In line with H1, morality-based groups cooperated more and kept fewer monetary units to themselves: d = -0.35, *95% CI* [-.55, -.15], F (1, 394) = 12, p = .001. In line with H2, they contributed more to the weak parochialism pool: d = 0.45, *95% CI* [.25, .64], F (1, 394) = 19.7, p < .001. We found supporting evidence for H3b, but not H3a: morality-based groups showed more universal cooperation: d = 0.25, *95% CI* [.05, .44], F (1, 394) = 6.2, p = .014, but less strong parochialism: d = -0.29, *95% CI* [-.49, -.10], F (1, 394) = 8.6, p = .005, than non-morality-based groups.

Figure 1

Contributions to Each Pool by Condition



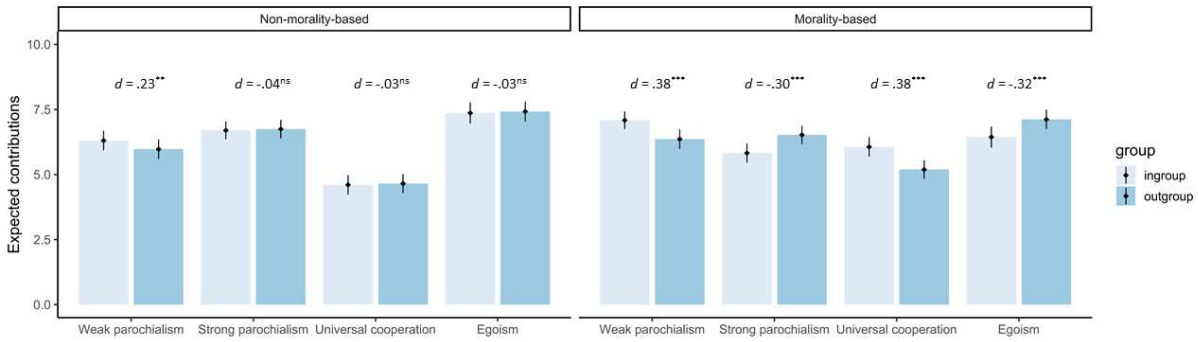
Note. The dots represent the means and error bars represent the 95% confidence intervals.

Exploratory analyses

Differences in trust and expectations. The expectations from the ingroup and outgroup members differed significantly in the morality condition, but not in the control condition (see effect sizes and *p*-values on Fig. 2). The morality-based groups expected the ingroup to contribute more to the weak parochialism and universal cooperation pools, but less to the strong parochialism and egoism pools. Essentially, participants in this condition expected the ingroup to be "moral" in a prosocial-universalistic sense: to help, but not to harm others. In contrast, there were no differences between expectations from the ingroup and the outgroup in the control condition, except for a small difference in expectations for weak parochialism.

Figure 2

Expected Contributions to Each Pool from Ingroup and Outgroup Members in Morality- and Non-Morality-Based Groups



Note. n.s. p > .05, ** p < .01, *** p < .001

Interestingly, most of the differences between the groups were driven by different expectations from the ingroup, not the outgroup. Morality-based groups anticipated more weak parochialism (d = .32, p < .001), less strong parochialism (d = -.36, p = .001), more universal cooperation (d = .58, p < .001), and less egoism (d = -.34, p = .001) from the ingroup than non-morality-based groups. The expectations from the outgroup did not differ in weak parochialism (d = .15, p = .092), strong parochialism (d = -.09, p = .390), and egoism (d = -.12, p = .250). The only difference was that morality-based groups anticipated slightly more universal cooperation from outgroups than non-morality-based groups (d = .22, p = .040).

The groups did not differ significantly on the general measure of trust towards the ingroup: M_m = 4.86, SD = 1.5, M_{nm} = 4.64, SD = 1.6, d = .14, p = .152, or the outgroup: M_m = 3.72, SD = 1.6, M_{nm} = 3.87, SD = 1.7, d = -.10, p = .344. However, the morality-based groups scored higher on the measure of parochial trust (the difference score between ingroup and outgroup trust): M_m = 1.14, SD = 1.5, M_{nm} = 0.76, SD = 1.4, d = .27, p = .008.

These findings point to a potential mediating role of trust in the effect of shared morality on intergroup cooperation. Since the study included multiple measures of trust (eight

expectations and three general measures of trust), we first explored which of these measures are both affected by condition and are related to behavior. Four of the trust measures were related to both condition and cooperative behavior: expectations of weak parochialism, strong parochialism, and universal cooperation from ingroup members, and the measure of parochial trust (see the results of path analysis in SOM, Table S4). We next conducted a path analysis in *lavaan* (Rosseel, 2012) to test whether these measures explain some of the variance in the link between condition and behavior.

Trust and expectations as mediators. Table 2 reports the results of the path analysis. We observed a clear pattern of relations between expectations and behavior: the more participants expected others to contribute to a certain pool, the more they themselves contributed to that pool, with similar effect sizes across the pools $\beta \approx .40$-$.50$. All three ingroup expectations mediated the link between the condition (morality-based group vs. control) and the behavior, but parochial trust did not. Note that these significant results are conditional upon the assumptions of the mediation model being correct (Fiedler et al., 2018).

Expectations of weak parochialism from ingroup members partially explained the positive link between morality-based group membership and weak parochialism (β_{ind} = .07**), although the main effect of condition remained significant (β = .13*). The negative effect of morality-based group membership on strong parochialism was partially explained by lower expectations of strong parochialism from ingroup members (β_{ind} = - .07**), and partially by higher expectations of weak parochialism (β_{ind} = -.02*) and universal cooperation (β_{ind} = -.05**) from ingroup members. The main effect of condition on strong parochialism after accounting for these indirect effects was not significant (β = -.06, p = .270). Finally, the positive effect of morality-based group membership on universal cooperation was partially explained by higher expectations of universal cooperation from ingroup members (β_{ind} = .14***). The main effect of condition after accounting for this indirect effect was not significant (β = .07, p = .374).

Table 2

Direct and indirect effects of condition on cooperative behavior

	b	*SE*	*z-value*	*p-value*	*β*
Weak parochialism (Pool A)					
Direct effects					
Condition: Morality-based group	0.71	0.29	2.50	.013	.13
Ingroup expectation: Pool A	0.45	0.06	8.17	< .001	.39
Ingroup expectation: Pool B	-0.11	0.06	-1.92	.055	-.09
Ingroup expectation: Pool C	-0.03	0.05	-0.61	.542	-.03
Parochial trust	0.21	0.09	2.24	.025	.11
Indirect effects					
via ingroup exp: Pool A	0.39	0.13	3.08	.002	.07
via ingroup exp: Pool B	0.09	0.06	1.67	.095	.02
via ingroup exp: Pool C	-0.05	0.08	-0.61	.545	-.01
via parochial trust	0.08	0.05	1.63	.104	.01
Total indirect	0.51	0.16	3.08	.002	.09
Strong parochialism (Pool B)					
Direct effects					
Condition: Morality-based group	-0.35	0.32	-1.10	.270	-.06
Ingroup expectation: Pool A	-0.16	0.06	-2.54	.011	-.12
Ingroup expectation: Pool B	0.49	0.06	8.08	< .001	.39
Ingroup expectation: Pool C	-0.22	0.06	-3.81	< .001	-.19
Parochial trust	0.11	0.10	1.05	.293	.05
Indirect effects					
via ingroup exp: Pool A	-0.13	0.07	-2.02	.044	-.02
via ingroup exp: Pool B	-0.43	0.14	-3.13	.002	-.07
via ingroup exp: Pool C	-0.33	0.10	-3.12	.002	-.05
via parochial trust	0.04	0.04	0.96	.336	.01
Total indirect	-0.85	0.19	-4.49	< .001	-.14
Universal cooperation (Pool C)					
Direct effects					
Condition: Morality-based group	0.29	0.33	0.89	.374	.04
Ingroup expectation: Pool A	-0.13	0.06	-2.00	.046	-.09
Ingroup expectation: Pool B	-0.10	0.06	-1.57	.118	-.07
Ingroup expectation: Pool C	0.65	0.06	10.68	< .001	.51
Parochial trust	-0.13	0.11	-1.16	.244	-.05
Indirect effects					
via ingroup exp: Pool A	-0.11	0.06	-1.71	.087	-.02
via ingroup exp: Pool B	0.09	0.06	1.42	.155	.01
via ingroup exp: Pool C	0.95	0.20	4.85	< .001	.14
via parochial trust	-0.05	0.04	-1.05	.296	-.01
Total indirect	0.88	0.22	4.03	< .001	.13

To conclude, we found no evidence that morality-based groups exhibit more hostility towards outgroups than non-morality-based groups, contradicting earlier evidence from studies with natural groups (Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). Instead, morality-based groups were more likely to prefer a non-parochial, mutually beneficial option, and expected less strong parochialism from ingroup members. We further provided some initial evidence on the role of trust, showing that only expectations of cooperation from the ingroup, but not the outgroup, factor into participants' decisions. These findings are consistent with the ingroup primacy assumption – the argument that prejudice and discrimination are rooted in expectations from and concern for the ingroup, not the desire to harm the outgroup (Brewer & Caporael, 2006). Having shared beliefs about what is right and what is wrong with ingroup but not outgroup members does not seem to be sufficient to cross the boundary between "ingroup love" to "outgroup hate".

Study 2: Registered Replication and Extension

Study 2 aimed to replicate the effect of morality-based vs. non-morality-based group membership on intergroup cooperation in a different cultural setting and to test the mediating effect of trust in this relationship. Based on our earlier findings, we expect morality-based groups to exhibit less egoism (H1), more weak parochialism (H2), and more universal cooperation (H3) than non-morality-based groups. Furthermore, we expect the effect of condition on contributions to be partially explained by expectations of contributions from ingroup members (H4). Specifically, we predict that the effect of condition on weak parochialism will be partially explained by higher expectations of weak parochialism from ingroup members (positive indirect effect, H4a). The effect of condition on strong parochialism will be partially explained by lower expectations of strong parochialism from ingroup members (negative indirect effect, H4b). The effect of condition on universal cooperation will be partially explained by higher expectations of universal cooperation from

ingroup members (positive indirect effect, H4c). The preregistration protocol is available at https://aspredicted.org/VZY_XRG.

Method

Sample

The study was approved by the ethics committee of the Psychology Department at [Anonymized] University. To test the replicability of our findings in a different cultural context, we recruited participants from the UK. Participants were recruited via the online platform Prolific. A total of 526 participants completed the study and 176 were excluded (see inclusion and exclusion criteria). The effective sample size is N=350 (49.7% men, 49.1% women, 1.1% non-binary; M_{age}=37.6, SD_{age}=12.1). All participants had at least secondary school education and 67% had a university degree.

Power analysis. We based our power analysis on the smallest effect of interest, which is the indirect effect of condition on weak parochialism via expectations of weak parochialism from the ingroup. We ran a power analysis using the *simsem* package in R (Pornprasertmanit et al., 2021). We used the dataset from Study 1 to extract parameter estimates. The power was estimated by running 1000 simulations at different sample sizes (300, 350, 400). All three indirect effects had a power of .90 or above at N = 350. We collected data until 350 valid (i.e., satisfying the inclusion criteria) responses were recorded.

Inclusion and exclusion criteria. Participants residing in the UK and over 18 years old could participate in the study. Data were excluded if (1) the participant did not make any contributions in the game or did not answer the comprehension questions (N=0 excluded), or (2) the participant incorrectly answered more than two comprehension questions for the game (N=176 excluded), or (3) the contributions made are not in the allowed range of 0-10 units (N=0 excluded).

Design and procedure

Manipulation. After providing informed consent, participants were randomly assigned to either the experimental or the control condition. As in Study 1, participants in the experimental condition filled out the 15-item Moral Foundations Questionnaire (MFQ). To make the control condition more comparable with the experimental condition, we made two changes compared to Study 1. First, since the MFQ can make moral issues salient and thus affect participants' cooperative behavior (e.g., increasing the likelihood of universal cooperation or decreasing the likelihood of strong parochialism), we included this questionnaire in the control condition as well, before the group assignment task. Second, to make the nature of the group assignment task more similar between the conditions, participants rated 15 works of art on a 6-point scale from 0 – "Don't like it at all" to 5 – "Like it very much", paralleling the structure of the MFQ questionnaire. Irrespective of participants' responses, they were assigned to a neutrally labelled group "X". In the experimental condition, they were informed that they will be making an economic decision in a group of four people who use similar criteria to judge what is right and what is wrong, with another group of four people (group Y), who use different criteria to judge what is right and what is wrong. In the control condition, the groups were defined by similarities and differences in preferences for art.

After completion of the group assignment task, participants filled out measures of group identification and perceived belief- and status-indicativeness of groups. They then played the Intergroup Parochial and Universal Cooperation (IPUC) game and filled out the remaining measures of expectations, perceived group entitativity and essentialism, trust, and attitudes. The game instructions were identical to Study 1 (English version adopted from Aaldering & Böhm, 2020) with one modification. Instead of lottery tickets, participants were playing for a bonus based on their performance in the game: 20% of all participants were randomly selected and received a compensation of £0.2 for each point earned in the game. All participants received £3 for their participation. The average bonus payment was £4.3.

Measures. The measures of ingroup and outgroup attitudes and entitativity were the same as in Study 1. We used a slightly modified version of the essentialism scale, replacing the informativeness item with an item about definiteness of group boundaries ("The boundaries between groups X and Y are clear-cut and definite"). Expectations of contributions from ingroup and outgroup members were measured on scales representing the amount of contributions (0 – 10) instead of the degree of certainty as in Study 1. General trust towards ingroup and outgroup members were measured by three items, which were averaged to form a single score: "I think members of [group] are trustworthy", "I trust members of [group] to make contributions that benefit others", and "I think I can rely on members of [group] in this decision-making task" on a scale from 1 – *"Strongly disagree"* to 5 – *"Strongly agree"*. We also included a measure of generalized trust as a potential control variable: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (WVS, 2015; Yamagishi & Yamagishi, 1994). Finally, we measured perceived morality, warmth, and competence of players who contributed to each of the pools for exploratory purposes (e.g., "Players who contributed most of their chips to pool A/B/C/D are…", with a list of traits adapted from Brambilla et al., 2011).

Quality checks. To check the effectiveness of the manipulation, we measured perceived belief- and status-indicativeness of groups as in Study 1. A successful manipulation should increase the perception of belief-indicativeness of the group in the experimental condition but have no effect on the perception of status-indicativeness. Additionally, we measured identification with the ingroup and the outgroup with two items each ("I see myself as a member of Group X/Y" and "I identify with Group X/Y"). We expect participants in both conditions to identify more with the ingroup than the outgroup. After receiving instructions for the IPUC game, participants answered six comprehension questions adopted from

(Aaldering & Böhm, 2020). Participants who incorrectly answered more than two of these comprehension questions were excluded from analysis.

Analysis Plan and Timeline

Data preprocessing. Observations that meet the exclusion criteria outlined above were excluded. All descriptive statistics were checked to ensure correctness of coding and absence of technical errors. Outliers were not excluded. We tested the reliability of all scales and did not identify any reliability issues (all $\alpha > .65$). We checked whether the manipulation was successful by estimating the differences in identification with the ingroup vs. outgroup using a paired sample t-test. Finally, we tested the effect of condition on measures of belief- and status-indicativeness of groups with one-way ANOVAs.

Main analysis. We ran a series of one-way ANOVAs, predicting contributions to each pool from condition. All *p*-values were adjusted for false discovery rate using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Cohen's *d* are used as an estimate of the effect size. Next, we planned to test a path model predicting contributions to the three pools (weak parochialism, strong parochialism, and universal cooperation) from condition (dummy coded: 0 – control, 1 – experimental), with three mediators: expectations of weak parochialism, strong parochialism, and universal cooperation from the ingroup.

Exploratory analysis. We tested whether the experimental and control groups differ by relevant demographic or psychological variables, such as gender, age, political affiliation, and generalized trust. If any significant differences are found, the analyses will be repeated including these variables as controls. We further report differences between the groups in identification, ingroup and outgroup attitudes and ingroup bias, perceptions of group entitativity and essentialism, and ingroup, outgroup, and parochial trust. Ingroup bias is calculated as the difference between ingroup and outgroup attitude, and parochial trust as the difference between ingroup and outgroup trust. Finally, since the MFQ questionnaire is

included in both conditions, we further explore the relationships between moral foundations and cooperative behavior.

Timeline. The Stage 1 registered report was accepted on April 8, 2022. The data was collected on April 14, 2022.

Open practices

All study materials, including data, materials, and code, are available via the Open Science Framework (OSF) platform: https://osf.io/uqwf3/?view_only=83a684d367464b3b94a6d39d5763551c. The links to the preregistration protocols are included in the Method section of each study.

Results and Discussion

The manipulation was successful. In both conditions, participants identified more with group X (the ingroup) than group Y (the outgroup; experimental: $d = 1.38$, t (167) = 17.98, $p < .001$, control: $d = 1.63$, t (181) = 21.94, $p < .001$). Participants in the morality condition perceived the group as more belief-indicative ($M = 2.63$, $SD = 0.99$) than participants in the control condition ($M = 1.47$, $SD = 0.67$; $d = 1.38$, F (1,348) = 166.3, $p < .001$). The groups were perceived as equally indicative of status (in both groups: $M = 1.3$, $SD = 0.6$; $d = 0.01$, F (1,348) = 0.01, $p = .908$).

Preregistered hypotheses

We found no significant differences in contributions between the experimental and control conditions (see Table 3). Descriptively, morality-based groups showed less egoism and less weak parochialism, but more universal cooperation.

Table 3

Contributions to each pool by morality- and non-morality-based groups

Pool	Morality-based		Non-morality-based		Cohen's *d*	*F*	*p*-value
	M	*SD*	*M*	*SD*			
A: weak parochialism	2.82	3.15	3.19	3.09	- 0.12	1.20	.550
B: strong parochialism	0.97	1.95	1.01	1.66	- 0.02	0.03	.855

C: universal cooperation	4.17	3.93	3.52	3.45	0.18	2.76	.389
D: egoism	2.04	3.02	2.29	3.16	- 0.08	0.59	.588

Note. All *p*-values are adjusted for false discovery rate using the Benjamini–Hochberg procedure

Since condition had no significant effect on contributions, we didn't test the mediation model as planned. Instead, we tested whether expectations from the ingroup and the outgroup predicted contributions in the combined sample. Table 4 presents the results.

Table 4

Predicting contributions from ingroup and outgroup expectations

	b	*SE*	*z-value*	*p-value*	*β*
Weak parochialism (Pool A)					
Ingroup expectation: Pool A	0.73	0.08	9.68	<.001	.55
Ingroup expectation: Pool B	-0.19	0.08	-2.44	.015	-.12
Ingroup expectation: Pool C	-0.46	0.07	-6.63	<.001	-.37
Ingroup expectation: Pool D	-0.17	0.07	-2.49	.013	-.15
Outgroup expectation: Pool A	-0.02	0.07	-0.24	.810	-.01
Outgroup expectation: Pool B	0.01	0.07	0.11	.911	.01
Outgroup expectation: Pool C	0.06	0.07	0.85	.395	.05
Outgroup expectation: Pool D	-0.02	0.06	-0.26	.795	-.02
Strong parochialism (Pool B)					
Ingroup expectation: Pool A	-0.12	0.05	-2.64	.008	-.17
Ingroup expectation: Pool B	0.52	0.05	11.15	<.001	.62
Ingroup expectation: Pool C	-0.13	0.04	-3.19	.001	-.20
Ingroup expectation: Pool D	-0.18	0.04	-4.57	<.001	-.31
Outgroup expectation: Pool A	-0.01	0.04	-0.33	.743	-.02
Outgroup expectation: Pool B	-0.09	0.04	-2.07	.039	-.12
Outgroup expectation: Pool C	-0.01	0.04	-0.34	.737	-.02
Outgroup expectation: Pool D	0.05	0.04	1.26	.206	.09
Universal cooperation (Pool C)					
Ingroup expectation: Pool A	-0.41	0.09	-4.51	<.001	-.27
Ingroup expectation: Pool B	-0.21	0.09	-2.22	.026	-.12
Ingroup expectation: Pool C	0.78	0.08	9.24	<.001	.54
Ingroup expectation: Pool D	-0.31	0.08	-3.82	<.001	-.24
Outgroup expectation: Pool A	0.03	0.09	0.34	.733	.02
Outgroup expectation: Pool B	0.06	0.09	0.73	.466	.04
Outgroup expectation: Pool C	0.03	0.08	0.42	.674	.03
Outgroup expectation: Pool D	0.06	0.07	0.78	.435	.05

As in Study 1, contributions were predicted by expectations from the ingroup, but not the outgroup. The more participants expected the ingroup members to contribute to a certain pool, the more they themselves contributed to that pool.

Exploratory analyses

We found no significant differences between conditions in any of the relevant demographic or psychological variables, such as gender, age, generalized trust, and religiosity (see SOM), suggesting that the randomization was successful.

Group perceptions. Table 5 presents mean differences in trust, attitudes, identity, and perceived group entitativity and essentialism. Although we observed no differences in cooperative behavior between the conditions, perceptions of the groups were quite different. Morality-based groups exhibited stronger ingroup bias in attitudes and trust. In both cases, the effects were driven mostly by the more positive perceptions of the ingroup rather than more negative perceptions of the outgroup. Morality-based groups were also seen as more entitative and essentialized.

Table 5

Differences between conditions in group perceptions

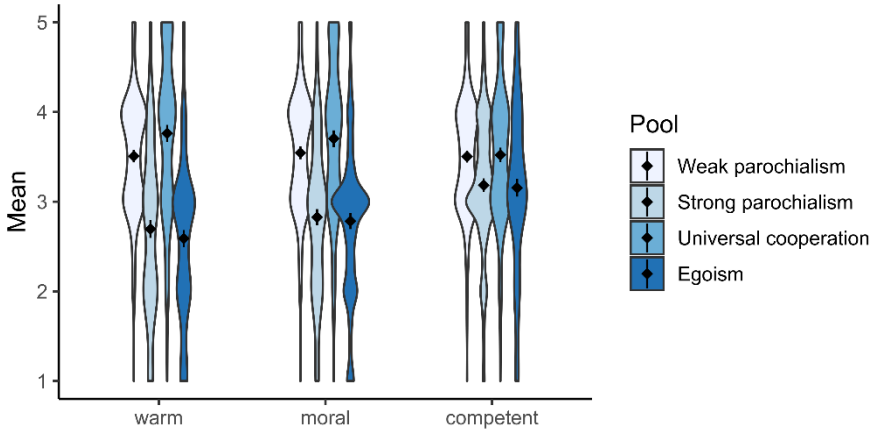
			Morality-based		Non-morality-based		*Cohen's d*	*p-value*
	min	*max*	*Mean*	*SD*	*Mean*	*SD*		
Ingroup identification	1	5	3.68	0.88	3.67	0.80	< 0.01	.926
Ingroup attitude	0	100	66.14	19.0	58.82	21.1	0.36	< .001
Outgroup attitude	0	100	47.52	16.3	49.4	19.8	0.10	.331
Ingroup bias	-19	80	18.62	19.1	9.18	13.6	0.57	< .001
Ingroup trust	1	5	3.58	0.67	3.20	0.73	0.54	< .001
Outgroup trust	1	5	2.90	0.69	3.05	0.71	-0.22	.038
Parochial trust	-1.7	4	0.68	0.89	0.15	0.48	0.76	< .001
Entitativity (outgroup)	1	5	3.78	0.82	3.20	0.95	0.66	< .001
Essentialism (outgroup)	1	5	2.66	0.82	2.32	0.85	0.40	< .001

Person perceptions. There were no differences between the conditions in how participants perceived others who contributed to each of the pools (see Table S5). As Figure 3 shows, participants in both conditions had a shared understanding that those who contributed

to the universal cooperation and weak parochialism pools were more warm, competent, and moral than those who contributed to the strong parochialism or egoism pools. Whereas contributors to the universal cooperation and weak parochialism pools were perceived as equally competent (t (348) = 0.44, p = .656), those who contributed to the universal cooperation pool were perceived as significantly warmer (t (348) = 4.42, p < .001) and more moral (t (348) = 2.90, p = .004) than those who contributed to the weak parochialism pool.

Figure 3

Perceptions of warmth, morality, and competence of persons contributing to each of the pools



Moral foundations. Since condition had no effect on contributions, we tested whether moral foundations were predictive of these contributions in the pooled sample. We ran a path model, predicting contributions to pools A, B, and C from the five moral foundations. All moral foundations combined explained only 2-5% of variance in contributions. Two moral foundations accounted for this explained variance: ingroup loyalty predicted less weak parochialism (β = -.20, p = .006) and more universal cooperation (β = .19, p = .011), and respect for authority predicted more weak parochialism (β = .28, p = .001) and less universal cooperation (β = -.28, p = .001). The full model is reported in Table S6 of the SOM.

To sum up, unlike in Study 1, we found no differences in cooperative intergroup behavior between morality and non-morality based groups. We did, however, find much stronger ingroup bias in attitudes and trust in morality-based vs. non-morality based groups.

How can this attitude-behavior discrepancy be explained? One potential explanation is that unlike in earlier studies, where participants were incentivized with lottery tickets, actual money were at stake here. This might have motivated participants to suppress their biases and choose the behavioral strategy that is most likely to give higher returns, which is the universal cooperation strategy. Participants in this study contributed more to the universal cooperation pool than in both previous studies.

To summarize the findings from the three studies, we conducted a mini meta-analysis (Goh et al., 2016) using the *metafor* (Viechtbauer, 2010) package in R. We calculated Hedge's *g* for each study. Overall, morality-based groups across the three studies showed less egoism (g = -0.26, *95%CI* [-.48, -.06], p = .013) and more universal cooperation (g = 0.22, *95%CI* [.09, .36], p = .001) than non-morality based groups. There were no significant differences in weak (g = 0.17, *95%CI* [-.17, .51], p = .329) or strong parochialism (g = -0.14, *95%CI* [-.33, .04], p = .136).

General discussion

When does "ingroup love" turn into "outgroup hate"? Previous studies conducted on natural groups suggest that centrality of morality to the group's identity is one such condition: morality-based groups showed more hostility towards outgroups than non-morality-based groups (Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). We set out to test this hypothesis in a minimal group setting, using the recently developed Intergroup Parochial and Universal Cooperation (IPUC) game.

Across three pre-registered studies, we found no evidence that morality-based groups show more hostility towards outgroups than non-morality-based groups. Instead, morality-based groups exhibited less egoism and more universal cooperation (helping both the ingroup and the outgroup) than non-morality-based groups. This finding is consistent with earlier research showing that salience of morality makes people more cooperative (Capraro et al., 2019). Importantly, our morality manipulation was not specific to any pro-cooperation moral

norm. Simply asking participants to think about the criteria they use to judge what is right and what is wrong was enough to increase universal cooperation.

Our findings are inconsistent with research showing stronger outgroup hostility in morality-based groups (Parker & Janoff-Bulman, 2013; Weisel & Böhm, 2015). The key difference between the set of studies presented here and the earlier studies that find outgroup hostility in morality-based groups is the use of natural groups in the latter. What potential confounding variables might account for the emergence of outgroup hostility in natural groups?

Our manipulation of morality-based group membership was maximally generic: participants were only told that the outgroup members "use different criteria to judge what is right and what is wrong". We did not specify which criteria exactly they use or how different they are from the one's used by the ingroup. The degree or the specific subject of difference or disagreement between the ingroup and the outgroup might be decisive for the emergence of outgroup hostility. Most real-life moral disagreements are (or are perceived to be) oppositional in nature: if one moral statement is true, the other is false, and it usually has serious real-life consequences. For example, let's take attitudes towards abortion (as in Parker & Janoff-Bulman, 2013). A world where most people believed abortions are morally unacceptable is incompatible with a world where most people believed they are morally acceptable. Therefore, this type of moral disagreement would be perceived as a zero-sum game: the world that you want to live in will not be possible if the opposite understanding of good and bad dominates. This would mean that previous findings of outgroup hostility in morality-based groups can be explained by perceptions of intergroup conflict. Besides intergroup conflict, other types of threats can be associated with morality-based group membership in natural groups. For example, absence of validation of one's fundamental moral beliefs can be perceived as a threat to one's worldview and a fear of being perceived as

bad or immoral by others can constitute an identity threat. In short, it is likely that for morality-based groups to show outgroup hostility, some type of threat must also be present.

Consistent with earlier findings (Bilancini et al., 2020), we found some evidence of stronger ingroup bias in morality-based groups: this was the case for behavior in Study 1 and attitudes and trust in Study 2. Importantly, these effects were driven exclusively by expectations from the ingroup and were unrelated to expectations from the outgroup in both studies. This finding provides additional evidence in support of the ingroup-focused explanation of prejudice and discrimination: negative attitudes towards outgroups are secondary to the concern for and desire to benefit the ingroup (Brewer & Caporael, 2006). Whereas ingroup favoritism is nearly universal and easy to induce (Balliet et al., 2014; Buhl, 1999), it seems that some type of threat is necessary for people to show hostility towards outgroups.

References

Aaldering, H., & Böhm, R. (2020). Parochial Versus Universal Cooperation: Introducing a Novel Economic Game of Within- and Between-Group Interaction. *Social Psychological and Personality Science*, *11*(1), 36–45. https://doi.org/10.1177/1948550619841627

Arvan, M. (2019). The dark side of morality: Group polarization and moral epistemology. *Philosophical Forum*, *50*(1), 87–115. https://doi.org/10.1111/phil.12213

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*(1), 289–300. https://www.jstor.org/stable/2346101

Bilancini, E., Boncinelli, L., Capraro, V., Celadin, T., & Di Paolo, R. (2020). "Do the right thing" for whom? An experiment on ingroup favouritism, group assorting and moral suasion. *Judgment and Decision Making*, *15*(2), 182–192. https://doi.org/10.2139/ssrn.3486398

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135–143. https://doi.org/10.1002/ejsp.744

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149–166. https://doi.org/10.1111/j.2044-

8309.2010.02011.x

Brandt, M. J., & Crawford, J. T. (2019). Studying a heterogeneous array of target groups can help us understand prejudice. *Current Directions in Psychological Science*, *28*(3), 292–298. https://doi.org/10.1177/0963721419830382

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B., & Caporael, L. R. (2006). An evolutionary perspective on social identity: Revisiting groups. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and Social Psychology* (pp. 143–161). Psychosocial Press. https://doi.org/10.4324/9780203782965

Buhl, T. (1999). Positive-negative asymmetry in social discrimination: Meta-analytical evidence. *Group Processes & Intergroup Relations*, *2*(1), 51–58. https://doi.org/https://doi.org/10.1177/1368430299021004

Buss, L. W. (1987). *The evolution of individuality*. Princeton University Press. http://www.jstor.org/stable/j.ctt7zvwtj

Capraro, V., Jagfeld, G., Klein, R., Mul, M., & de Pol, I. van. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-48094-4

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate?: Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, *60*(1), 47–69. https://doi.org/10.1086/701478

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In *Nebraska Symposium on Motivation, 1962.* (pp. 275–320). Univer. Nebraska Press.

Ellemers, N., & van den Bos, K. (2012). Morality in Groups: On the Social-Regulatory Functions of Right and Wrong. *Social and Personality Psychology Compass*, *6*(12), 878–889. https://doi.org/10.1111/spc3.12001

Ellemers, N., & Van der Toorn, J. (2015). Groups as moral anchors. *Current Opinion in Psychology*, *6*, 189–194. https://doi.org/10.1016/j.copsyc.2015.08.018

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*(November 2017), 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. https://doi.org/10.1037/a0034726

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In *Advances in Experimental Social Psychology* (1st ed., Vol. 47). Copyright © 2013, Elsevier Inc. All rights reserved. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

https://doi.org/10.1037/a0021847

Grigoryan, L. (2020a). Crossed categorization outside the lab: Findings from a factorial survey experiment. *European Journal of Social Psychology*, *50*(5), 983–1000. https://doi.org/10.1002/ejsp.2656

Grigoryan, L. (2020b). Perceived similarity in multiple categorisation. *Applied Psychology*, *69*(4), 1122–1144. https://doi.org/10.1111/apps.12202

Grigoryan, L., Cohrs, J. C., Boehnke, K., van de Vijver, F. A. J. R., & Easterbrook, M. J. (2022). Multiple categorization and intergroup bias: Examining the generalizability of three theories of intergroup relations. *Journal of Personality and Social Psychology*, *122*(1), 34–52. https://doi.org/10.1037/pspi0000342

Grigoryan, L., Jones, B. H., Cohrs, J. C., Boehnke, K., & Easterbrook, M. J. (2022). Differentiating between belief-indicative and status-indicative groups improves predictions of intergroup attitudes. *Personality and Social Psychology Bulletin*. OnlineFirst. https://doi.org/10.1177/01461672221092852

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, *33*, 1–36. https://doi.org/https://doi.org/10.1111/j.1559-1816.2003.tb02071.x

Halevy, N., Bornstein, G., & Sagiv, L. (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm: Research article. *Psychological Science*, *19*(4), 405–411. https://doi.org/10.1111/j.1467-9280.2008.02100.x

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kouzakova, M., Ellemers, N., Harinck, F., & Scheepers, D. (2012). The Implications of Value Conflict: How Disagreement on Values Affects Self-Involvement and Perceived Common Ground. *Personality and Social Psychology Bulletin*, *38*(6), 798–807. https://doi.org/10.1177/0146167211436320

Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The Problem with Morality: Impeding Progress and Increasing Divides. In *Social Issues and Policy Review* (Vol. 12, Issue 1). https://doi.org/10.1111/sipr.12045

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, *127*(2), 187–208. https://doi.org/10.1037/0033-2909.127.2.187

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234–249. https://doi.org/10.1037/0022-3514.93.2.234

Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *Journal of Personality and Social Psychology*, *90*(4), 629–643. https://doi.org/10.1037/0022-3514.90.4.629

Mummendey, A., & Otten, S. (1998). Positive–Negative Asymmetry in Social Discrimination. *European Review of Social Psychology*, *9*(1), 107–143.

https://doi.org/10.1080/14792779843000063

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love." *Social Justice Research*, *26*(1), 81–96. https://doi.org/10.1007/s11211-012-0175-6

Pornprasertmanit, S., Miller, P., Schoemann, A., Jorgensen, T. D., & Quick, C. (2021). *SIMulated Structural Equation Modeling*. https://simsem.org/

Qualtrics. (2005). *No Title* (No. 2015). Qualtrics. https://www.qualtrics.com

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12702–12707. https://doi.org/10.1073/pnas.1712921114

Rosenbaum, M. E. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, *51*(6), 1156–1166. https://doi.org/10.1037//0022-3514.51.6.1156

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, *28*(10), 1419–1429. https://doi.org/10.1177/014616702236873

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.1103/PhysRevB.91.121108

Weisel, O., & Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, *60*, 110–120.

https://doi.org/10.1016/j.jesp.2015.04.008

West, S. A., Mouden, C. El, & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, *32*(4), 231–262. https://doi.org/10.1016/j.evolhumbehav.2010.08.001

Wit, A. P., & Kerr, N. L. (2002). "Me versus just us versus us all" categorization and cooperation in nested social dilemmas. *Journal of Personality and Social Psychology*, *83*(3), 616–637. https://doi.org/10.1037/0022-3514.83.3.616

WVS. (2015). World Values Survey Wave 6 2010-2014. In *World Values Survey Association*.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 130–166.

Supplementary Online Materials

for

"Helping the ingroup versus harming the outgroup: Evidence from morality-based groups"

Pilot Study

Group identification was measured with four items (α = .85; example: "I identify with other members of the group") evaluated on a 7-point scale adopted from Doosje et al. (1995). Ingroup and outgroup attitudes were measured with one item ("What is your general attitude regarding the [target group]") adopted from Weisel & Böhm (2015). We measured perceptions of threat from the outgroup with ten items adopted from Cottrell & Neuberg (2005) and Parker & Janoff-Bulman (2013) (α = .95; example: "[target group] are a threat to my personal values"). Group entitativity was measured with four items capturing perceptions of groupness, similarity, interactivity, and common goals (α = .76; example: "[outgroup] qualifies as a group", adapted from (Blanchard et al., 2020; Lickel et al., 2000) and essentialism with three items capturing perceptions of immutability, stability, and informativeness (α = .33[1]; example: "[outgroup] is difficult to join or leave", adapted from (Bastian & Haslam, 2006; Demoulin et al., 2006). Table S1 presents the mean scores of these measures by condition.

Table S1

Identity, Attitude, Entitativity and Essentialism by Condition (Pilot Study)

			Morality-based		Non-morality-based		*Cohen's d*	*p-value*
	min	*max*	*Mean*	*SD*	*Mean*	*SD*		
Identity	1	7	3.63	1.4	3.64	1.1	<.001	.997
Ingroup attitude	1	7	4.60	1.0	4.44	0.9	0.17	.329
Outgroup attitude	1	7	4.18	0.8	4.05	0.7	0.17	.335

[1] Excluding the immutability item would increase reliability to α = .38. Since this change is not substantial, we decided to keep all three items.

Ingroup bias (ingroup – outgroup attitude)	-3	4	0.42	1.4	0.39	1.1	0.02	.888
Threat (outgroup)	1	9	3.18	1.9	2.59	1.6	0.33	.062
Entitativity (outgroup)	1	7	4.66	1.1	4.15	1.0	0.49	.007
Essentialism (outgroup)	1	7	3.54	1.0	3.46	1.1	0.08	.652

Study 1

We measured group identification (α = .88), perceived entitativity (α = .75) and essentialism (α = .63) same as in the pilot study. In- and outgroup attitudes were measured with a 100-degrees feelings thermometers, and ingroup bias as the difference between ingroup and outgroup evaluations.

Table S2

Identity, Attitude, Entitativity and Essentialism by Condition (Study 1)

			Morality-based		Non-morality-based		*Cohen's d*	*p-value*
	min	*max*	*Mean*	*SD*	*Mean*	*SD*		
Identity	1	7	3.95	1.57	3.94	1.54	< 0.01	.954
Ingroup attitude	0	100	58.6	23.7	55.0	26.5	0.14	.166
Outgroup attitude	0	100	40.1	20.9	40.4	23.4	-0.01	.898
Ingroup bias	-30	90	18.4	20.9	14.6	17.9	0.20	.052
Entitativity (outgroup)	1	7	4.53	1.03	4.37	1.19	0.15	.150
Essentialism (outgroup)	1	7	3.23	1.10	2.54	1.16	0.61	< .001

Table S3

Pattern of Contributions in Morality-Based and Non-Morality-Based Groups (Study 1)

			Morality-based			Non-morality-based		
	min	*max*	*Mean*	*SD*	*Med.*	*Mean*	*SD*	*Med.*
Weak parochialism (Pool A)	0	10	2.88	3.04	2.5	1.66	2.43	0
Strong parochialism (Pool B)	0	10	1.80	2.87	0	2.72	3.35	1
Universal cooperation (Pool C)	0	10	2.93	3.51	2	2.10	3.22	0
Egoism (Pool D)	0	10	2.38	2.99	1.5	3.51	3.45	2

Table S4

Path Model Predicting Contributions to Each Pool from Expectations and Parochial Trust

	b	*SE*	*z-value*	*p-value*	*β*
Pool A					
Ingroup expectation: Pool A	***0.558***	***0.09***	***6.223***	***< .001***	***.488***
Ingroup expectation: Pool B	-0.156	0.08	-1.885	.059	-.136
Ingroup expectation: Pool C	0.071	0.08	0.840	.401	.066
Ingroup expectation: Pool D	-0.074	0.08	-0.879	.380	-.073
Outgroup expectation: Pool A	-0.119	0.09	-1.382	.167	-.108
Outgroup expectation: Pool B	0.070	0.08	0.840	.401	.061
Outgroup expectation: Pool C	-0.162	0.08	-1.894	.058	-.140
Outgroup expectation: Pool D	-0.036	0.09	-0.409	.682	-.033
Parochial trust	***0.208***	***0.09***	***2.204***	***.028***	***.107***
Pool B					
Ingroup expectation: Pool A	-0.100	0.10	-1.016	.310	-.077
Ingroup expectation: Pool B	***0.359***	***0.09***	***3.954***	***< .001***	***.275***
Ingroup expectation: Pool C	***-0.314***	***0.09***	***-3.393***	***.001***	***-.258***
Ingroup expectation: Pool D	-0.061	0.09	-0.663	.507	-.053
Outgroup expectation: Pool A	-0.104	0.10	-1.095	.274	-.083
Outgroup expectation: Pool B	**0.203**	**0.09**	**2.207**	**.027**	**.153**
Outgroup expectation: Pool C	0.062	0.09	0.661	.508	.047
Outgroup expectation: Pool D	-0.089	0.10	-0.927	.354	-.071
Parochial trust	0.054	0.10	0.525	.600	.024
Pool C					
Ingroup expectation: Pool A	***-0.325***	***0.10***	***-3.302***	***.001***	***-.235***
Ingroup expectation: Pool B	-0.070	0.09	-0.774	.439	-.051
Ingroup expectation: Pool C	***0.652***	***0.09***	***7.058***	***< .001***	***.507***
Ingroup expectation: Pool D	-0.180	0.09	-1.95	.051	-.147
Outgroup expectation: Pool A	**0.260**	**0.09**	**2.744**	**.006**	**.196**
Outgroup expectation: Pool B	-0.104	0.09	-1.133	.257	-.074

Outgroup expectation: Pool C	-0.068	0.09	-0.720	.472	-.049
Outgroup expectation: Pool D	-0.145	0.09	-1.518	.129	-.110
Parochial trust	-0.147	0.10	-1.418	.156	-.062

Note. Measures that were affected by condition are in italic (see section "Differences in Trust" in the main text). Significant effects on contributions are in bold. Measures that were both affected by condition and predicted contributions are in bold and italic.

Study 2

The experimental and control conditions did not differ by gender ($\chi 2$ (1) = 0.05, p = .825), age (F (1,347) = 0.11, p = .738), generalized trust (F (1,348) = 0.07, p = .790), or religiosity (F (1,348) = 0.56, p = .456).

Table S5

Perceptions of persons contributing to each of the pools by condition and in the pooled sample

	Pooled sample		Morality-based		Non-morality-based		*F*	*p*
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*		
1. Pool A: warm	3.51	0.66	3.49	0.68	3.53	0.65	0.28	.597
2. Pool A: moral	3.54	0.67	3.51	0.69	3.58	0.65	0.78	.378
3. Pool A: competent	3.50	0.64	3.51	0.62	3.50	0.66	0.03	.852
4. Pool B: warm	2.69	0.95	2.65	0.99	2.73	0.90	0.66	.415
5. Pool B: moral	2.83	0.86	2.76	0.91	2.89	0.81	1.83	.177
6. Pool B: competent	3.18	0.72	3.17	0.72	3.20	0.73	0.14	.707
7. Pool C: warm	3.76	0.90	3.76	0.88	3.76	0.92	< .001	.995
8. Pool C: moral	3.70	0.88	3.73	0.87	3.68	0.89	0.37	.541
9. Pool C: competent	3.52	0.77	3.50	0.81	3.54	0.74	0.27	.604
10. Pool D: warm	2.59	0.90	2.61	0.90	2.57	0.90	0.18	.673
11. Pool D: moral	2.78	0.84	2.78	0.87	2.79	0.82	0.02	.880
12. Pool D: competent	3.15	0.93	3.09	0.93	3.22	0.93	1.78	.183

Table S6

Predicting contributions to each pool from moral foundations

	b	*SE*	*z-value*	*p-value*	*β*
Pool A					
Harm	0.08	0.27	0.31	.755	.022
Fairness	-0.02	0.27	-0.09	.932	-.006
Ingroup loyalty	-0.62	0.23	-2.74	.006	-.205
Respect for authority	0.88	0.26	3.41	.001	.284
Purity	-0.11	0.22	-0.49	.624	-.036
Pool B					
Harm	0.02	0.16	0.14	.891	.010
Fairness	-0.12	0.16	-0.75	.452	-.052
Ingroup loyalty	0.03	0.13	0.24	.811	.018
Respect for authority	0.20	0.15	1.33	.185	.111
Purity	0.10	0.13	0.81	.418	.059
Pool C					
Harm	0.31	0.31	0.98	.329	.067
Fairness	0.30	0.32	0.94	.348	.064
Ingroup loyalty	0.67	0.27	2.54	.011	.189
Respect for authority	-1.01	0.30	-3.34	.001	-.276
Purity	-0.03	0.25	-0.13	.901	-.009