

“Do the right thing” for whom? An experiment on ingroup favouritism, group assorting and moral suasion

Ennio Bilancini* Leonardo Boncinelli† Valerio Capraro‡ Tatiana Celadin§ Roberto Di Paolo¶

Abstract

In this paper we investigate the effect of moral suasion on ingroup favouritism. We report a well-powered, pre-registered, two-stage 2x2 mixed-design experiment. In the first stage, groups are formed on the basis of how participants answer a set of questions, concerning non-morally relevant issues in one treatment (assorting on non-moral preferences), and morally relevant issues in another treatment (assorting on moral preferences). In the second stage, participants choose how to split a given amount of money between participants of their own group and participants of the other group, first in the baseline setting and then in a setting where they are told to do what they believe to be morally right (moral suasion). Our main results are: (i) in the baseline, participants tend to favour their own group to a greater extent when groups are assorted according to moral preferences, compared to when they are assorted according to non-moral preferences; (ii) the net effect of moral suasion is to decrease ingroup favouritism, but there is also a non-negligible proportion of participants for whom moral suasion increases ingroup favouritism; (iii) the effect of moral suasion is substantially stable across group assorting and four pre-registered individual characteristics (gender, political orientation, religiosity, pro-life vs pro-choice ethical convictions).

Keywords: moral suasion, ingroup favouritism, dictator game, moral preferences.

1 Introduction

In the past years, behavioural scientists have provided converging evidence that social behaviour in economic games is not solely motivated by the monetary consequences of the available actions, but it is also motivated by moral preferences for doing what people think to be the right thing, beyond the monetary consequences that this action brings about. For example, Krupka and Weber (2013) found that people tend to be more altruistic in the dictator game in the “take frame”, compared to the dictator game in the “give frame”, in spite of the fact that these two frames give rise to monetarily equivalent decision problems. Crucially, this framing effect could be explained by preferences for making the socially appropriate choice: people tend to rate “taking

from the recipient” to be more socially inappropriate than “not giving to the recipient”. A conceptually similar result has been obtained for six dictator game frames in Capraro and Vanzo (2019), where the authors found that their framing effects can be explained by a change in the perception of what people think is the morally right thing to do. The fact that frames can affect choices by activating moral preferences is not limited to altruistic behaviour in the dictator game. For example, Eriksson et al. (2017) reported that moral considerations explain framing effects in ultimatum game rejections, whereas Capraro and Rand (2018) demonstrated that moral preferences explain framing effects in trade-off games pitting equity against efficiency, and drive not only altruistic behaviour in the dictator game, but also cooperative behaviour in the prisoner’s dilemma. Besides the empirical work, moral preferences have also been formalized in several economic models (Bénabou & Tirole, 2006; Levitt & List, 2007; López-Pérez, 2008; Andreoni & Bernheim, 2009; Della Vigna et al., 2012; Kessler & Leider, 2012; Alger & Weibull, 2013; Krupka & Weber, 2013; Kimbrough & Vostroknutov, 2016). And besides behavioural economics, the fact that at least altruistic giving is driven by morality has been highlighted by several scholars, to the point that fairness is considered to be one of foundations of morality (Haidt & Joseph, 2004; Graham et al., 2013; Haidt, 2012).

One of the main applications of this literature on moral preferences is the work on moral suasion. The idea is simple: if behaviour is driven by moral preferences, then making morality salient should affect behaviour. A number of works have provided support for this hypothesis. An earlier

This research was financially supported by IMT School for Advanced Studies Lucca through the PAI project PRO.CO.P.E. (Prosociality, Cognition and Peer Effects). The experiment was designed and conducted during Valerio Capraro’s visit to IMT School as a PRO.CO.P.E. guest scholar.

Copyright: © 2020. The authors license this article under the terms of the <http://creativecommons.org/licenses/by/3.0/> Creative Commons Attribution 3.0 License.

*IMT School of Advanced Studies, Lucca. Email: ennio.bilancini@imtlucca.it.

†Department of Economics and Management, University of Florence. Email: leonardo.boncinelli@unifi.it.

‡Corresponding author: Department of Economics, Middlesex University London. Email: v.capraro@mdx.ac.uk.

§IMT School of Advanced Studies, Lucca. Email: tatiana.celadin@imtlucca.it.

¶Department of Economics, University of Alicante; IMT School of Advanced Studies, Lucca. Email: roberto.dipaolo@imtlucca.it.

paper by Brañas-Garza (2007) found that telling dictators that "the other person relies on you" increases dictator game donations towards that person. A subsequent work by Dal Bó and Dal Bó (2014) reported that reminders of the Golden Rule increases cooperation in an iterated prisoner's dilemma. More recently, Capraro et al. (2019) showed that asking participants "What do you personally think is the morally right thing to do?", prior to making their decision, increases dictator game altruism and prisoner's dilemma cooperation. Scholars have also started applying moral suasion to redistribution decisions that have consequences outside the laboratory. Capraro et al. (2019) found that the aforementioned moral nudge increases online charitable donations to humanitarian organizations by 44%. Bott et al. (2019) reported that sending moral letters to tax payers decreases tax evasion. In sum, moral suasion is emerging as a useful tool to nudge prosocial behaviour both in and outside the lab.¹

Here, we extend this line of literature by studying the effect of moral suasion in dictator games in which dictators can favour a member of their own group, at a cost to a member of another group. This is an important case for two reasons.

One is practical. Scholars have been studying ingroup favouritism (and its companion, outgroup derogation) for decades, with the underlying motivation that this is what ultimately generates some of the lightest and, at the same time, some of the darkest expressions of human nature, democracy and civil rights, on the one hand, genocides and ethnic cleansings, on the other hand (Tajfel et al., 1971; Wilson, 1975; Nowak, 2006; Puurtinen & Mappes, 2008; Rusch, 2014). Therefore, exploring whether and how moral suasion affects ingroup favouritism may have significant practical implications.

The second reason is that theoretical predictions are, a priori, not obvious, as we argue below.

On the one hand, ingroup favouritism is considered to be one of the most fundamental behavioural bias among humans (Haidt, 2012; Baron et al., 2013; Curry et al., 2019). The seminal work by Tajfel et al. (1971) found that people discriminate between groups even when groups are assorted through a random, irrelevant categorization (see also Tajfel (1970); Tajfel (1974); Tajfel (1982)). Since then, ingroup favouritism has been observed in several economic contexts, ranging from public goods games (Krupp et al., 2008), dictator games (Whitt & Wilson, 2007; Chen & Li, 2009), charitable donation games (Pavey et al., 2011), ultimatum games (McLeish & Oxoby, 2011; Kubota et al., 2013), prisoner's dilemmas (Ahmed, 2007), and response games (Chen & Li, 2009), among many others (Everett et al., 2015). Ingroup favouritism is so widespread that some psychologists have

come to suggest that its psychological basis, *group identity*, ultimately descends from the uniquely human awareness of the inevitability of death: when people identify with a group, they embed themselves into something greater, that ultimately outlives the individual, and this allows the individual to reach the so-called "symbolic immortality" (Becker, 2007; Harmon-Jones et al., 1996; Arndt et al., 1997). Important for our work is that ingroup favouritism is widespread also as a moral rule: "help your group" has been recently found to be a universal moral rule across 60 societies (Curry et al., 2019). This is consistent with the morality-as-cooperation theory (Curry, 2016), as well as with moral foundations theory, according to which ingroup favouritism represents one of the fundamental dimensions of morality (Graham et al., 2013; Haidt, 2012; Haidt & Joseph, 2004).

Therefore, this line of literature suggests that moral suasion might increase ingroup favouritism, by making it salient that the right thing to do is to help your group.

Alternatively, it is possible that the effect of moral suasion on ingroup favouritism is not domain-general, but it depends on specific factors such as the decision context and the individual characteristics of the decision maker. In this paper, we focus on one contextual factor and four individual characteristics.

Among the contextual factors, the way the groups are assorted is likely to affect the strength of ingroup favouritism and, ultimately, the effect of moral suasion. For example, if people are grouped according to characteristics that are central to their identity, it is likely that they would display more ingroup bias, compared to situations in which they are grouped according to less central characteristics. Thus, moral suasion could increase ingroup favouritism in cases in which participants are assorted according to characteristics that generate a strong group identity, while leaving it unaffected, or even decreasing it, when participants are assorted according to characteristics that generate a weaker group identity. What characteristics are central to people's identity? Previous work has shown that people consider their moral traits to be the most essential part of their identity, even more so than emotional and autobiographical memory (Strohlinger & Nichols, 2014). Additionally, morality is often the first characteristic that people use to form impressions and evaluate others (Goodwin et al., 2014; Wojciszke et al., 1998). This suggests that it is possible that assorting participants according to their moral preferences would generate a stronger group identity, compared to when participants are assorted according to non-moral preferences; and this difference in group identity would translate into a difference in ingroup favouritism and the effect of moral suasion on ingroup favouritism.

To select individual characteristics, we took a pragmatic approach. Since we planned to conduct our experiment on Prolific and since Prolific allows experimenters to download some individual characteristics of the participants, we

¹A related stream of literature uses descriptive and injunctive norms to promote prosocial behaviour in the laboratory (Bicchieri & Xiao, 2009; Krupka & Weber, 2009; Zafar, 2011; Raihani & McAuliffe, 2014; D'Adda et al., 2017) and in the field (Frey & Meier, 2004; Croson et al., 2010; Cialdini et al., 1991; Ferraro & Price, 2013; Agerström et al., 2016; Goldstein et al., 2008; Hallsworth et al., 2017).

checked, prior to the experiment, the list of all the individual characteristics that were available for download and we selected those that we believed to be potentially relevant. In doing so, we selected four characteristics: gender, political orientation, religiosity, and pro-life vs. pro-choice convictions. We chose gender because previous research has found that males tend to display greater levels of social dominance orientation than females do (Sidanius et al., 1994), and social dominance orientation is known to be predictive of various forms of outgroup derogation, including racism (Sidanius & Pratto, 1993). A related line of work suggests that between-groups competition might be evolved especially among men, because it allowed men to gain access to mates and other resources (Vugt et al., 2007; McDonald et al., 2012). Therefore, it is possible that men display more ingroup favoritism and this could be even strengthened by moral suasion. We chose political orientation because previous work suggests the existence of a “prejudice gap” between conservatives and liberals (Chambers et al., 2013), with conservatives being more intolerant towards outgroups than liberals (see Sibley & Duckitt, 2008, for a meta-analysis). Consequently, it is possible that moral suasion decreases ingroup favoritism for liberals, while increasing it for conservatives. For similar reasons, we chose religiosity and pro-life vs. pro-choice convictions, being typically correlated with political conservatism (Malika et al., 2012). See also Enke et al. (2019).

Following this line of thoughts, we designed, pre-registered, and conducted a well-powered ($N=502$), 2×2 mixed-design experiment, in which the first, between-subjects, factor represents the way the groups are assorted (according to moral vs non-moral preferences), while the second, within-subject, factor represents the way people are asked to make dictator game decisions between individuals in their own group and individuals in the other group (baseline vs under moral suasion). We chose to use a within subject design for the dictator game decisions in order to be able to categorize participants in three types: those who, in response to moral suasion, discriminate more between groups; those who, in response to moral suasion, discriminate less between groups; those who, in response to moral suasion, do not change their strategy. This subdivision in types can help us shed light on the heterogeneity on people’s moral preferences.²

In a nutshell, our main results are: (i) participants tend to favour their own group to a greater extent when groups are assorted according to moral preferences, compared to

when they are assorted according to non-moral preferences; (ii) the net effect of moral suasion is to decrease ingroup favoritism, but there is also a non-negligible proportion of participants for whom moral suasion increases ingroup favoritism; (iii) the effect of moral suasion is substantially stable across group assorting and the four individual characteristics under consideration.

2 Method

We conducted an online 2×2 experiment on Prolific (Palan & Schitter, 2018), implemented in oTree (Chen et al., 2016). We recruited 502 participants living in the US at the time of the experiment. In the first stage of the experiment, participants were randomly assigned to one of two treatments designed to group together individuals with similar stated preferences. In one treatment, preferences are collected as answers to questions on issues that are morally relevant; in the other treatment, preferences are collected as answers to questions on issues that are not morally relevant. In the second stage, each participant played a randomized sequence of three variants of the dictator game (DG) in two distinct settings: the baseline setting and the moral suasion setting. The DG is a non-strategic game, where the decision-maker, the *dictator*, has to decide how to split a certain amount of money between herself and the *receiver*. In our study, the decision maker has to decide how to split 100 points (in 10-points increments) between a given pair of recipients. In the next subsections, we describe the experiment in more details. Full experimental instructions are reported in Appendix A.

2.1 Stage 1: Group Formation

In the first stage of the experiment, participants were randomly assigned to one of the two treatments: the “moral assorting” (250 participants) and the “non-moral assorting” (252 participants) treatments. In the moral assorting treatment, participants answered five questions concerning moral issues (see Appendix A) and had to indicate if they believe these issues are morally acceptable, morally wrong or if they have no opinion. In the non-moral assorting treatment, participants were asked their preferences on five non-moral issues (Appendix A). For each subject, her *own group* was defined for her as the set of participants who answered in the same way as she did to at least three out of the five questions, with the remaining participants forming the *other group*.

2.2 Stage 2: Dictator Games

In the second stage of the experiment, participants played DGs in two distinct settings with fixed order: first in the “baseline” setting and then in the “moral suasion” setting. In the baseline, each participant had to decide how to split 100

²Implementing a within-subject design comes also with some costs. On the one hand, the effect of moral suasion might decrease, because a within-subject design introduces demand for consistency (Samuelson & Zeckhauser, 1988); on the other hand, the effect of moral suasion might increase, due to experimenter demand effect. However, we believe this last issue to be less relevant in our case, because “demand effects refer to changes in behaviour due to cues about what constitutes appropriate behaviour” (Zizzo, 2010); therefore, people who change donation because of demand effect are still following a norm.

points in three different randomized DGs. In the first DG (which we call "DG own"), each participant had to divide points between herself and a randomly picked member of her own group. In the second DG (which we call "DG other"), each participant had to divide points between herself and a randomly picked member of the other group. In the third DG (which we call "DG own-other"), each participant had to divide points between a randomly picked member of her own group and a randomly picked member of the other group. After this, participants played in the moral suasion setting. Here, participants faced the same three DGs described above, but before allocating points they were told: "do what you think is morally right".

Lastly, participants were asked comprehension questions. We refer to Appendix A for full experimental instructions.

After collecting all the data, participants were randomly assigned to the role of decision-maker or receiver. Then one of the DGs was randomly selected and subjects were paired according to the selected DG. Participants received a payment according to their role in that particular DG. This payment methodology implies that all decisions have an impact on both the decision maker and the recipient. On average, participants gained 0.61 GBP, including the show-up fee (0.40 GBP).

2.3 Measures of Ingroup Favouritism

We operationalize ingroup favouritism through two different individual-level measures. One measure is constructed starting from the DGs in which the decision-maker is affected by her decision (the "DG own" and the "DG other"). Specifically, this measure is computed as the difference between how much a dictator gives to a randomly picked member of her own group and how much she gives to a randomly picked member of the other group. We call this measure *costly ingroup favouritism*. Note that this is different from the notion of parochial altruism, which requires the action to be (i) costly to the decision-maker, (ii) beneficial for the decision maker's ingroup, and (iii) costly for outgroup members, all at the same time (Böhm et al., 2018; Choi & Bowles, 2007). In particular, costly ingroup favouritism does not necessarily involve harming outgroup members; it might be the case that a dictator still gives some amount to outgroup members, but less compared to ingroup members. We also consider a measure of *costless ingroup favouritism*, whereby helping one's own group does not cost anything to the decision maker. This measure is constructed starting from the DG in which the dictator is not affected by her decision ("DG own-other"). Specifically, this measure is computed as the difference between 50 (the equal split) and how much a dictator gives to a randomly picked member of the other group.

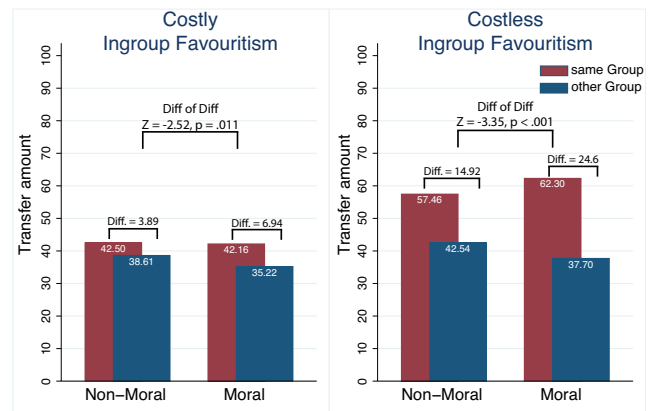


FIGURE 1: **Ingroup favouritism is stronger when assorting is based on moral preferences, compared to when it is based on non-moral preferences.** The average of the costly ingroup favouritism measure in the moral treatment ($M=6.94$, $SD=12.27$) is 3.05 points higher than it is in the non-moral one ($M=3.89$, $SD=16.57$) (left chart). The average of the costless ingroup favouritism measure in the moral treatment ($M=12.3$, $SD=20.37$) is 4.83 points higher than it is in the non-moral one ($M=7.46$, $SD=15.69$) (right chart).

2.4 Research questions

We pre-registered three research questions:

1. Does assorting based on moral preferences generate more ingroup favouritism than assorting based on non-moral preferences?
2. Does moral suasion mitigate ingroup favouritism?
3. Does moral suasion affect ingroup favouritism differently when group assorting is based on moral preferences compared to when it is based on non-moral preferences?

Furthermore, we pre-registered that we would test the role of: gender, political orientation, religiosity, pro-life vs pro-choice ethical convictions. The pre-registration is available at: <https://aspredicted.org/k4r34.pdf>.

3 Results

Our first research question is whether assorting based on moral preferences generates more ingroup favouritism compared to the case where assorting is based on non-moral preferences. Figure 1 suggests that, on average, the answer is positive for both measures of ingroup favouritism. Wilcoxon rank-sum test confirms this finding. When assorting is based on moral preferences, the average of the costly ingroup favouritism measure is 3.05 points greater than it is when assorting is based on non-moral preferences, $Z=2.52$, $p = .011$ (left chart). A similar result holds for the costless

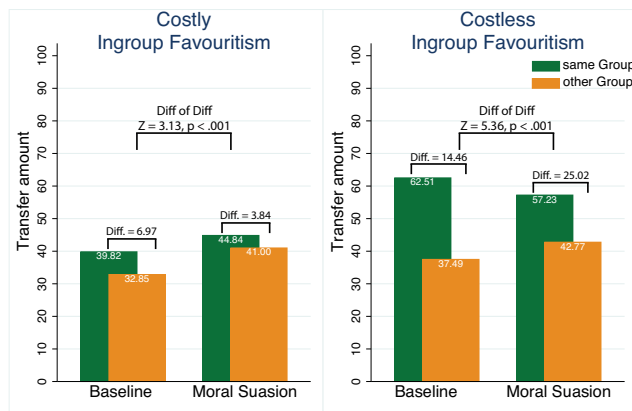


FIGURE 2: **Ingroup favouritism is weaker under moral suasion compared to the baseline.** The average of the costly ingroup favouritism measure under moral suasion (M=3.84, SD= 12.95) is 3.13 points lower than it is in the baseline (M=6.97, SD= 16.02) (left chart). The average of the costless ingroup favouritism measure under moral suasion (M=7.23, SD=16.22) is 5.28 points lower than it is in the baseline (M=12.51, SD=19.89) (right chart).

ingroup favouritism measure. In the moral treatment, the average of the costly ingroup favouritism measure is 4.83 points greater than it is in the non-moral treatment, $Z=3.35$, $p<.001$ (right chart).

The second research question is whether moral suasion mitigates ingroup favouritism compared to the baseline. Figure 2 suggests that, on average, the answer is positive for both measures of ingroup favouritism. Wilcoxon rank-sum test confirms this finding. In the baseline, the average of the costly ingroup favouritism measure is 3.13 points higher than it is under moral suasion $Z=3.13$, $p<.001$ (left chart). A similar result holds for the costless ingroup favouritism measure. In the baseline, the average of the costless ingroup favouritism measure is 5.28 points higher than it is under moral suasion $Z(502)=5.37$, $p<.001$ (right chart). Splitting the effect of moral suasion by DG decision, we find that moral suasion increases prosociality both in “DG own” and “DG other” (linear regression: $t = 4.47$, $p < .001$; $t = 4.61$, $p < .001$). However, the increase of prosociality in “DG other” is even greater ($t = 6.40$, $p < .001$), and this is ultimately the reason why, on average, moral suasion attenuates ingroup favouritism.

The fact that, on average, moral suasion decreases ingroup favouritism does not imply that moral suasion has the effect of decreasing ingroup favouritism for all the participants. In fact, we now use the within-subject factor to show that there are three non-negligible types of participants, which characteristically differ on how they react to moral suasion. In the first type we include participants who increased their ingroup favouritism under moral suasion (*persuaded parochialists*: ingroup favouritism under moral suasion is

TABLE 1: Frequencies of the different types using costly and costless ingroup favouritism measures, across treatments.

Treatment	Costly		Costless	
	Non-Moral	Moral	Non-Moral	Moral
Persuaded parochialists	8.3%	8.4%	6.3%	6.0%
Persuaded universalists	19.8%	16.0%	24.6%	21.6%
Unpersuaded	71.8%	75.6%	69.1%	72.4%

Note. *Persuaded parochialists*: participants who increase ingroup favouritism under moral suasion, compared to the baseline. *Persuaded universalists*: participants who decrease ingroup favouritism under moral suasion, compared to the baseline. *Unpersuaded*: participants who do not change ingroup favouritism under moral suasion, compared to the baseline.

higher than ingroup favouritism in the baseline.³) In the second type we classify participants who decrease their ingroup favouritism under moral suasion (*persuaded universalists*: ingroup favouritism under moral suasion is lower than ingroup favouritism in the baseline). In the third type we categorize participants who do not change behaviour under moral suasion (*unpersuaded*: ingroup favouritism under moral suasion is equal to ingroup favouritism in the baseline). Table 1 reports the proportions of these different behavioural types in each treatment and using both the costly and costless ingroup favouritism measures. Across treatments and measures, the majority of participants is unpersuaded (on average, 72.2%), a substantial proportion is persuaded universalist (on average, 20.5%), and, interestingly, a smaller but non-negligible proportion is persuaded parochialist (on average, 7.3%). To provide further evidence that the apparent existence of persuaded parochialist is not the result of random variation, we distinguish persuaded parochialists in the costly setting from those of the costless setting. In particular, we construct two variables: one takes value 1 if a participant is classified as persuaded parochialist according to the costly ingroup favouritism measure, and 0 otherwise; and the other is analogously defined by using the costless ingroup favouritism measure. Pearson’s Chi-squared test shows that these two variables are not independent ($p < .001$). Since these measures have been collected separately, this suggests that they are sensitive to the same causal factor.

We now move to the third research question, whether there is any difference in how moral suasion affects in-

³In this case we use the word *parochialism*, because persuaded parochialists actually act in a way that is costly to outgroup members, compared to the baseline. Therefore, all three properties listed by Böhm et al. (2018) that classify parochial decisions are satisfied.

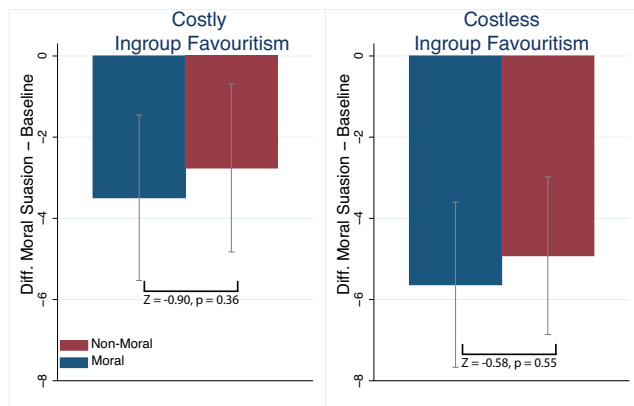


FIGURE 3: The effect of moral suasion on ingroup favouritism is substantially stable across group assorting. The reduction in the costly ingroup favouritism measure due to moral suasion in the moral treatment is not significantly different from its reduction in the non-moral treatment (left chart, $p = .36$). The reduction in the costless ingroup favouritism measure due to moral suasion in the moral treatment is not significantly different from its reduction in the non-moral treatment (right chart, $p = .55$). Error bars represent 95% CI.

group favouritism in the two treatments, i.e., when assorting is based on moral preferences and when it is based on non-moral ones. Figure 3 reports the difference in ingroup favouritism under moral suasion and the baseline, across treatments and measures. Mixed-design ANOVA predicting ingroup favouritism as a function of group assorting, moral suasion, and their interaction, shows no significant interaction, both for the costly ingroup favouritism measure ($F = 0.25, p = 0.620$) and for the costless one ($F = 0.25, p = 0.617$). This result is confirmed if we use the rank-sum test. Using both measures of ingroup favouritism, we find that the effect of moral suasion is substantially stable across group assorting (costly measure: $Z = .90, p = .36$; costless measure: $Z = .59, p = .55$). Table 2 reports the fractions of participants for whom the ingroup favouritism measure is strictly greater than 0, across measures and treatments. Again we find that the effect of moral suasion is substantially stable across treatments, as confirmed by mixed-design ANOVA, which reports a non-significant interaction between group assorting and moral suasion (costly measure: $F = 1.96, p = 0.1618$; costless measure: $F = 0.95, p = 0.3297$).

Finally, we ask whether the differential effect of moral suasion depends on the four pre-registered individual characteristics that we could download from Prolific: gender, political orientation, religiosity, and pro-life vs pro-choice ethical convictions. To do so we coded 4 dummy variables: *gender*, 1 if female and 0 if male; *liberal*, 1 if Liberal and 0 if Conservative/Moderate; *no_religious*, 1 if agnostic,

atheist or no religious, and 0 if religious; *prochoice*, 1 if a participant reports to have pro-choice ethical convictions and 0 if a participant reports to have pro-life ethical convictions. Table 3 reports average characteristics of participants by treatments and types. The type subdivision is substantially stable across treatments and individual characteristics. Females, liberals, no religious and pro-choice people tend to be slightly more represented among persuaded universalists than they are among persuaded parochialists. However, the only statistically significant difference is that persuaded universalists are less religious than persuaded parochialists, $Z(61) = 2.69, p = .007$. Therefore, we conclude that the effect of moral suasion is substantially stable across the four individual characteristics that we considered.

We also tested the effect of the four individual characteristics on baseline levels of ingroup favouritism. For each individual characteristic, we used a mixed-design ANOVA predicting ingroup favouritism as a function of assorting, moral suasion, the individual characteristics, and all their two- and three-way interactions (with all variables centered on their mean). The three-way interaction is significant only in the case of *liberal* and only for the costly ingroup favouritism measure ($F = 5.86, p = 0.016$). Instead, for the two-way interactions, only the one between moral suasion and *religiosity* turns out to be marginally significant effect for both measures of ingroup favouritism (costly: $F = 3.37, p = 0.067$; costless: $F = 3.53, p = 0.061$). All other interactions are not statistically significant (all p 's > 0.1).

4 Discussion

The interest in moral suasion stems, at least in part, from being a cheap and possibly effective policy tool that could be applied to foster prosocial behaviours. While the literature on moral behaviour has so far produced a substantial body of evidence showing the effectiveness of moral suasion, its dependence on the identity of the recipients of the decision-maker's actions is far less studied, leaving open the possibility that individuals react to moral suasion by reducing prosociality towards some *types* of recipients. This paper has addressed this issue in the setting of a decision to split a given amount of money between members of one's own group and members of another group, providing experimental evidence that, *on average*, moral suasion increases pro-sociality towards both the ingroup and the outgroup; however, the increase towards the outgroup is greater than the increase towards the ingroup, and this results in the fact that ingroup favouritism, on average, declines under moral suasion.

This effect exists when groups are defined in terms of similarity with respect to answers to both non-morally relevant questions and morally relevant questions, even if, in the latter case, the initial level of ingroup favouritism is higher. We

TABLE 2: Percentages of participants for which the ingroup favouritism measure is strictly greater than 0, across treatments and ingroup favouritism measures.

Treatment	Costly			Costless		
	Baseline	Moral Suasion	Difference	Baseline	Moral Suasion	Difference
Non-Moral	24.9%	11.3%	13.6%	36.7%	17.5%	19.2%
Moral	27.7%	20.0%	7.7%	43.5%	28.6%	14.9%

TABLE 3: Average of different groups’ characteristics in the non-moral and moral treatments.

Treatment	Persuaded parochialist		Persuaded universalist		Unpersuaded	
	Non-Moral	Moral	Non-Moral	Moral	Non-Moral	Moral
Gender	0.55	0.52	0.60	0.65	0.59	0.53
Liberal	0.53	0.37	0.43	0.58	0.58	0.59
No Religious	0.62	0.43	0.70	0.78	0.65	0.62
Pro Choice	0.83	0.76	0.84	0.86	0.82	0.75

Note. *Gender*: 1 if female; 0 if male. *Liberal*: 1 if Liberal; 0 if Conservative/Moderate. *No religious*: 1 if agnostic, atheist or no religion; 0 if religious. *Pro Choice*: 1 if pro-choice ethical convictions; 0 if pro-life ethical convictions.

would like to stress that we are not suggesting that moral assorting is *qualitatively* different from non-moral assorting. The fact that moral assorting generates stronger initial ingroup favouritism compared to non-moral assorting is likely to be driven by the fact that moral assorting makes group identity *quantitatively* stronger.

What is important to note, instead, is that the negative effect of moral suasion on ingroup favouritism holds *on average*. When we look at how participants change their decision in response to moral suasion, we find a small, but non-negligible, group of participants who increase their level of ingroup favouritism. Finally, the effect of moral suasion appears to be substantially stable across four (pre-registered) individual characteristics: gender, political orientation, religiosity, pro-choice vs pro-life ethical convictions. The only significant effect is that persuaded universalists tend to be less religious than persuaded parochialists. These findings have potential applications outside the laboratory, as they suggest that making the morality of an action salient might be a practical and effective tool for decreasing ingroup favouritism, on average, while also having the drawback of actually increasing ingroup favouritism for a non-negligible proportion of participants.

This study also relates to research exploring whether moral assorting affects ingroup favouritism. Parker & Janoff-Bulman (2013) showed that when people are assorted according to their preferences on abortion, then they report feeling stronger positive ingroup emotions and negative out-

group emotions, compared to when they are assorted according to whether they prefer the Red Sox or the Yankees (this study was conducted in Massachusetts, where the rivalry between these two teams is particularly strong). Weisel & Böhm (2015) divided people in groups according to their preferences about football clubs vs. political parties and found that people actively harm outgroup members only in the case in which assorting is based on political preferences. Our work is conceptually in line with this literature as our first result shows that people display greater ingroup favouritism when they are assorted according to moral preferences, compared to when they are assorted according to non-moral preferences.

This is the first study investigating the effect of moral suasion on ingroup favouritism. As such, it does have several limitations that might suggest directions for future research. First, we focused on only one contextual factor that might impact the effect of moral suasion on ingroup favouritism: assorting according to moral preferences. Future work should explore how moral suasion affects ingroup favouritism when group assorting is based on other characteristics that are likely to activate a strong group identity. For example, some scholars argue that ingroup favouritism is rooted in our evolutionary tribal past (Wilson, 1975; Nowak, 2006; Puurtinen & Mappes, 2008; Rusch, 2014; Fu et al., 2012; Masuda & Fu, 2015). This suggests that characteristics that provide group advantages (e.g., language) might be better candidates than characteristics that primarily provide individual advantages

(e.g., skill specialization). Such an investigation could lead to identify specific cases where moral suasion is particularly effective, and possibly others where it delivers undesirable effects.

Another potential variable of interest is the mode of cognition – whether decisions are made under “system 1” (often referred to as *intuition*) or “system 2” (often referred to as *deliberation*) (Evans & Stanovich, 2013). One study found that promoting intuition favours ingroup favouritism (De Dreu et al., 2015); two more found that promoting intuition favours cooperation with outgroup members (Rand et al., 2015; Everett et al., 2017). See Capraro (2019) for a review. Consequently, it is possible that the effect of moral suasion on ingroup favouritism is moderated, or even mediated, by the mode of cognition. Future work should explore this possibility.

A similar limitation regards individual characteristics. In this work we focused on gender, political orientation, religiosity, and pro-life vs pro-choice ethical convictions. The results suggest that the effect of moral suasion is substantially stable across these characteristics. Future work should explore the role of other personal characteristics. Another limitation regards the fact that our results are based on a laboratory experiment. Since moral suasion potentially represents a very practical tool for policy interventions, a key direction for future work is to test the external validity of our findings.

In sum, we studied the effect of moral suasion on ingroup favouritism. Our main results are: (i) in the baseline, participants tend to favour their own group to a greater extent when groups are assorted according to moral preferences, compared to when they are assorted according to non-moral preferences; (ii) the net effect of moral suasion is to decrease ingroup favouritism, but there is also a non-negligible proportion of participants for whom moral suasion increases ingroup favouritism; (iii) the effect of moral suasion is substantially stable across group assorting and four pre-registered individual characteristics (gender, political orientation, religiosity, pro-life vs. pro-choice ethical convictions). Future work should test the effect of other contextual factors and individual characteristics.

References

- Agerström, J., Carlsson, R., Nicklasson, L., & Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52, 147–153.
- Ahmed, A. M. (2007). Group identity, social distance and intergroup bias. *Journal of Economic Psychology*, 28(3), 324–337.
- Alger, I. & Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6), 2269–2302.
- Andreoni, J. & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607–1636.
- Arndt, J., Greenberg, J., Solomon, S., Pyszczynski, T., & Simon, L. (1997). Suppression, accessibility of death-related thoughts, and cultural worldview defense: Exploring the psychodynamics of terror management. *Journal of Personality and Social Psychology*, 73(1), 5.
- Baron, J., Ritov, I., & Greene, J. D. (2013). The duty to support nationalistic policies. *Journal of Behavioral Decision Making*, 26(2), 128–138.
- Becker, E. (2007). *The denial of death*. Simon and Schuster.
- Bénabou, R. & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5), 1652–1678.
- Bicchieri, C. & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191–208.
- Böhm, R., Rusch, H., & Baron, J. (2018). The psychology of intergroup conflict: a review of theories and measures. *Journal of Economic Behavior & Organization*.
- Bott, K. M., Cappelen, A. W., Sorensen, E., & Tungodden, B. (2019). You’ve got mail: a randomised field experiment on tax evasion. *Management Science*.
- Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4), 477–486.
- Capraro, V. (2019). The dual-process approach to human sociality: A review. Available at SSRN 3409146.
- Capraro, V., Jagfeld, G., Klein, R., Mul, M., & van de Pol, I. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports*, 9, 11880.
- Capraro, V. & Rand, D. G. (2018). Do the Right Thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*, 13, 99–111.
- Capraro, V. & Vanzo, A. (2019). The power of moral words: Loaded language generates framing effects in the extreme dictator game. *Judgment and Decision Making*, 14, 309–317.
- Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science*, 24(2), 140–149.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
- Chen, Y. & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1), 431–57.
- Choi, J.-K. & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, 318(5850), 636–640.
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior.

- In *Advances in experimental social psychology*, volume 24 (pp. 201–234). Elsevier.
- Crosno, R. T., Handy, F., & Shang, J. (2010). Gendered giving: The influence of social norms on the donation behavior of men and women. *International Journal of Nonprofit and Voluntary Sector Marketing*, 15(2), 199–213.
- Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In *The evolution of morality* (pp. 27–51). Springer.
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate. *Current Anthropology*, 60(1), 47–69.
- Dal Bó, E. & Dal Bó, P. (2014). "Do the right thing:" The effects of moral suasion on cooperation. *Journal of Public Economics*, 117, 28–38.
- De Dreu, C. K., Dussel, D. B., & Velden, F. S. T. (2015). In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. *Frontiers in psychology*, 6, 572.
- Della Vigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, 127(1), 1–56.
- D'Adda, G., Capraro, V., & Tavoni, M. (2017). Push, don't nudge: Behavioral spillovers and policy instruments. *Economics Letters*, 154, 92–95.
- Enke, B., Rodríguez-Padilla, R., & Zimmermann, F. (2019). Moral Universalism: Measurement and Heterogeneity. Available at https://www.ifo.de/DocDL/cesifo1_wp7921.pdf.
- Eriksson, K., Strimling, P., Andersson, P. A., & Lindholm, T. (2017). Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction. *Journal of Experimental Social Psychology*, 69, 59–64.
- Evans, J. S. B. & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241.
- Everett, J. A., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, 9, 15.
- Everett, J. A., Ingbretsen, Z., Cushman, F., & Cikara, M. (2017). Deliberation erodes cooperative behavior—Even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *Journal of Experimental Social Psychology*, 73, 76–81.
- Ferraro, P. J. & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1), 64–73.
- Frey, B. S. & Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review*, 94(5), 1717–1722.
- Fu, F., Tarnita, C. E., Christakis, N. A., Wang, L., Rand, D. G., & Nowak, M. A. (2012). Evolution of in-group favoritism. *Scientific Reports*, 2, 460.
- Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3), 472–482.
- 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.
- 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*, volume 47 (pp. 55–130). Elsevier.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Haidt, J. & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55–66.
- Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14–31.
- Harmon-Jones, E., Greenberg, J., Solomon, S., & Simon, L. (1996). The effects of mortality salience on intergroup bias between minimal groups. *European Journal of Social Psychology*, 26(4), 677–681.
- Kessler, J. B. & Leider, S. (2012). Norms and contracting. *Management Science*, 58(1), 62–77.
- Kimbrough, E. O. & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608–638.
- Krupka, E. & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, 30(3), 307–320.
- 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.
- Krupp, D. B., Debruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior*, 29(1), 49–55.
- Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychological Science*, 24(12), 2498–2504.
- Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174.
- López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior*, 64(1), 237–267.

- Malka, A., Lelkes, Y., Srivastava, S., Cohen, A. B., & Miller, D. T. (2012). The association of religiosity and political conservatism: The role of political engagement. *Political Psychology, 33*(2), 275–299.
- Masuda, N. & Fu, F. (2015). Evolutionary models of in-group favoritism. *F1000prime reports, 7*.
- McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1589), 670–679.
- McLeish, K. N. & Oxoby, R. J. (2011). Social interactions and the salience of social identity. *Journal of Economic Psychology, 32*(1), 172–178.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*(5805), 1560–1563.
- Palan, S. & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.
- 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.
- Pavey, L., Greitemeyer, T., & Sparks, P. (2011). Highlighting relatedness promotes prosocial motives and behavior. *Personality and Social Psychology Bulletin, 37*(7), 905–917.
- Puurtinen, M. & Mappes, T. (2008). Between-group competition and human cooperation. *Proceedings of the Royal Society B: Biological Sciences, 276*(1655), 355–360.
- Raihani, N. J. & McAuliffe, K. (2014). Dictator game giving: The importance of descriptive versus injunctive norms. *PloS ONE, 9*(12), e113826.
- Rand, D. G., Newman, G. E., & Wurzbacher, O. M. (2015). Social context and the dynamics of cooperative choice. *Journal of Behavioral Decision Making, 28*(2), 159–166.
- Rusch, H. (2014). The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences, 281*(1794), 20141539.
- Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty, 1*(1), 7–59.
- Sibley, C. G. & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review, 12*(3), 248–279.
- Sidanius, J. & Pratto, F. (1993). Racism and Support of Free-Market Capitalism: A Cross-Cultural Analysis. *Political Psychology, 14*(3), 381–401.
- Sidanius, J., Pratto, F., & Bobo, L. (1994). Social dominance orientation and the political psychology of gender: A case of invariance? *Journal of Personality and Social Psychology, 67*(6), 998–1011.
- Strohming, N. & Nichols, S. (2014). The essential moral self. *Cognition, 131*(1), 159–171.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*(5), 96–103.
- Tajfel, H. (1974). Social identity and intergroup behaviour. *Information (International Social Science Council), 13*(2), 65–93.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*(1), 1–39.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178.
- Vugt, M. v., Cremer, D. D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis. *Psychological Science, 18*(1), 19–23.
- Weisel, O. & Böhm, R. (2015). “Ingroup love” and “outgroup hate” in intergroup conflict between natural groups. *Journal of experimental social psychology, 60*, 110–120.
- Whitt, S. & Wilson, R. K. (2007). The dictator game, fairness and ethnicity in postwar Bosnia. *American Journal of Political Science, 51*(3), 655–668.
- Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences, 72*(1), 143–146.
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251–1263.
- Zafar, B. (2011). An experimental investigation of why individuals conform. *European Economic Review, 55*(6), 774–798.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98.

Appendix: Experimental instructions

Group Formation

Participants were randomly divided into two treatments. We report the instructions for both treatments.

Assorting based on Moral preferences:

Below you see a list of issues. For each one of them, regardless of whether or not you think it should be legal, please indicate whether you personally believe that in general it is morally acceptable, morally wrong or if you have no opinion:

- Abortion: Morally Acceptable/Morally Wrong/No Opinion;
- Doctor assisted suicide: Morally Acceptable/Morally Wrong/No Opinion;

- Death penalty: Morally Acceptable/Morally Wrong/No Opinion;
- Gay or lesbian relations: Morally Acceptable/Morally Wrong/No Opinion;
- Prostitution: Morally Acceptable/Morally Wrong/No Opinion.

Assorting based on Non-Moral preferences:

Below you see a list of questions. For each one of them, please indicate which option you personally prefer, or if you have no opinion.

- Where do you prefer to go during vacation? Sea/Mountain/No Opinion;
- Where do you prefer to watch movies? Movie Theater/Home/No Opinion;
- Where do you prefer to do physical activity? Gym/Outdoor/No Opinion;
- Which social network do you prefer? Instagram/Facebook/No Opinion;
- Which animal do you prefer? Dog/Cat/No Opinion.

Here participants moved to the next screen.

Two groups will be formed, YOUR GROUP and the OTHER GROUP, using the answers that you and the other participants have given so far. YOUR GROUP is formed by you and by those participants with answers most similar to yours. OTHER GROUP is formed by the remaining participants, those with answers least similar to yours.

Ingroup Favouritism

Here participants faced three randomized DG in the baseline setting.

In the next screens, you will make a number of decisions about how to divide 100 Points between two participants, drawn from either YOUR GROUP, the OTHER GROUP or both. Once the survey is over, payments will be determined according to either one of the decisions you made or one of the decisions made by another participant that involves you. 100 Points correspond to payment of 0.50 GBP.

- You have to allocate Points between YOU and another member of YOUR GROUP.
- You have to allocate Points between YOU and a member of the OTHER GROUP.
- You have to allocate Points between a member of YOUR GROUP (not you) and a member of the OTHER GROUP.

Here all participants faced three randomized DG in the moral suasion setting.

In the next decisions, you have to decide how to divide Points according to what you think is morally right.

- You have to allocate Points between YOU and another member of YOUR GROUP. Do what you think is morally right.
- You have to allocate Points between YOU and a member of the OTHER GROUP. Do what you think is morally right.
- You have to allocate Points between a member of YOUR GROUP (not you) and a member of the OTHER GROUP. Do what you think is morally right.

Here the comprehension questions.

- What is the decision that lets you obtain the highest payment?
- What is the decision that lets the other participant obtain the highest payment?
- What is the decision that lets you and the other participant obtain the same payment?