# The rationale of in-group favoritism: An experimental test of three explanations*

Giuseppe Ciccarone
*Department and Economics and Law, Sapienza University of Rome*
giuseppe.ciccarone@uniroma1.it

Giovanni Di Bartolomeo
*Department and Economics and Law, Sapienza University of Rome; Board of the Experts, Ministry of Economy and Finance, Italy*
giovanni.dibartolomeo@uniroma1.it

Stefano Papa
*Department and Economics and Law, Sapienza University of Rome*
stefano.papa@uniroma1.it

This version: 19 June 2019

**Abstract.** We aim to empirically investigate the rationale of in-group-favoritism. As potential explanations, we concentrate on intrinsic preferences for own-group members and motivations related to individuals caring about the beliefs of others (founded on guilt aversion). We also consider their intersection. Our evidence shows that in-group-favoritism cannot be accounted for by changes in expectations. This suggests that preferences *per se* are the most powerful explanation of social identity.

***JEL* codes:** A13, C91, D03, D64, D90.

**Keywords:** Social identity, in-group-favoritism, second-order beliefs, guilt aversion, causation.

Corresponding author's address: Prof. Stefano Papa, Department of Economics and Law, Faculty of Economics, Sapienza University of Rome, via del Castro Laurenziano 9, 00191, Rome, Italy Phone +39 06 49766400 - Fax +39 06 4957606.

* The authors are grateful for comments to Giuseppe Attanasi, Pierpaolo Battigalli, Maria Bigoni, Gary Charness, Stefano DellaVigna, Daniela Di Cagno, Martin Dufwenberg, Werner Guth, John Hey, Rachel Kranton, Robert Neugeboren, Matteo Rizzoli, Marie Claire Villeval, Paolo Zaffaroni and the participants in the 3rd Workshop on Psychological Game Theory on "Emotions and Economic Theory," organized by the *Society for the Advancement of Behavioral Economics*, in the 12th International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World (LUISS, Rome), in the 1st CoCoLab Workshop (Nice, France). A previous version circulated with the title "Is social identity belief independent?" The views expressed in this paper are those of the authors and should not be attributed to the Ministry of Economy and Finance.

## 1. Introduction

Social Identity Theory (SIT) predicts that intergroup behavior depends on perceived group status differences, the perceived legitimacy and stability of these differences, and the perceived ability to move from one group to another (Tajfel, 1970, 2010; Tajfel and Turner, 1979, 1986). In-group favoritism is a common feature of such behavior, as individuals often endorse resource distributions that maximize the positive distinctiveness of an in-group, in contrast to an out-group, at the expense of personal self-interest.[1] The aim of this paper is to add to the literature on social identity, expectations and belief manipulation, by focusing on the underlying motivations of behavior leading to in-group favoritism.

Both preferences and beliefs have been singled out as important elements to formalize and explain SIT (Everett *et al*., 2015). Two main viewpoints can be identified. On the one hand, Akerlof and Kranton (2000), Chen and Li (2009), and Chen and Chen (2011) incorporate social identity into the social preference model and suggest that people have an intrinsic preference for group members, i.e., the deviation from group identity causes disutility.[2] On the other hand, Güth *et al.* (2009), Ockenfels and Werner (2014), and Daskalova (2018) explore motivations related to group members' beliefs when individuals care about the others' beliefs (second-order beliefs), which are consistent with guilt aversion.[3] According to this view, as individuals care about the others' beliefs, social identify may shape beliefs and thus affect decisions.

We can accordingly group potential motivations consistent with SIT under three main labels.[4]

1. Bob has an intrinsic (or *per se*) preference for group members. Therefore, Bob will favor Ann if she is a member of the same group, independently of Bob's beliefs about Ann's beliefs. It is worth noting that, assuming asymmetric information, Bob favors Ann even if Ann does not know the identity of Bob. What matters is Bob's information about Ann's membership. We refer to this idea as *Group Identity-based Explanation* (*GIE*).
2. If Bob dislikes letting the others' expectations down (i.e., he is guilt averse), in-group favoritism can emerge as the equilibrium of a psychological game where social identity shapes beliefs. Formally, it occurs if Bob believes that the other members believe that he will behave in a more

[1] See, among others, Chen and Li (2009), Guth *et al*. (2009), Chen and Chen (2011), Guala *et al*. (2013), Ockenfels and Werner (2014). Section 2 briefly discusses the economic literature on social identity.

[2] This view moves along the lines of the moral commitment issue broadly discussed in the promise keeping literature (see, e.g., Ellingsen and Johannesson, 2004; Vanberg, 2008; Ellingsen *et al*., 2010).

[3] The concept of guilt aversion was popularized by Charness and Dufwenberg (2006) and further developed by Battigalli and Dufwenberg (2007). See also Attanasi *et al*. (2016). Empirical support to guilt aversion is provided, among others, by Bellemare *et al*. (2011) and Khalmetski *et al*. (2015). For a survey, see Attanasi and Nagel (2008).

[4] For the sake of exposition, we refer to Bob as the person who favors, and we refer to Ann as the person who is favored.

altruistic way with individuals belonging to his group than with people belonging to other groups. Assuming asymmetric information, in such a case, if Ann erroneously believes that Bob is a member of her group, Bob's beliefs on Ann's beliefs is that she expects to interact with a generous partner belonging to her group. Bob will then be generous with Ann even if she does not belong to his group.[5] We refer to this idea as *Beliefs Driven Explanation* (*BDE*).

3. Assuming that Bob's guilt sensitivity with respect to Ann depends on whether or not Ann is in Bob's group, then Bob dislikes letting Ann's expectations down only if she belongs to his group. In this case, like in *BDE*, in-group favoritism emerges as the equilibrium of a psychological game. However, unlike in *BDE*, considering again asymmetric information, Bob will no longer be generous with Ann if she erroneously believes that Bob is a member of her group. We refer to this idea as *Belief-Mediated-group-identity Explanation* (*BME*), which captures some elements form both *GIE* and *BDE*.

Understanding the rationale of in-group bias is important. In fact, although, in-group favoritism may offer a positive sense of belonging and affiliation (e.g., favoring your own sport team), often it can foster, rather than heal, conflicts, leading to destructive and hurtful behaviors (e.g., think about gossiping against others, bullying and scapegoating, pressuring group members to do what they individually do not feel comfortable in doing). On a much larger scale, social identity implies discrimination against people who do not share your race, ethnicity, religion, or nationality. Policies designed to attenuate its effects (or to promote it when they are positive) can hence benefit from disentangling the potential explanations. For instance, if beliefs mattered for in-group favoritism, discrimination could arise even without tastes or stereotypes against particular groups. In such a case, in order to eradicate discrimination, it would not be sufficient to eradicate the discriminants' preferences against diversity, but it would also be necessary to eradicate expectations of in-group favoritism.

The attempts to establish the relative importance of intrinsic preference vs. beliefs is not trivial, as motivations based on both are consistent with the observed empirical correlation between in-group favoritism and second-order beliefs. On the one hand, beliefs depend on preferences, as in forming their beliefs, or their beliefs on the others' beliefs, people try to predict the actions of the others, which depend on preferences. On the other hand, beliefs may affect preferences when individuals care about the others' beliefs. The empirical unraveling of the effects of beliefs and intrinsic preference is therefore made difficult by an endogeneity problem. In order to overcome this problem, we borrow from the literature on promise-

[5] Note that Bob's second-order beliefs, in fact, depend on what he knows Ann knows and not on what he knows.

keeping motivations[6] to develop an experimental design that can disentangle the effects driven by second-order beliefs from those stemming from intrinsic preference.

Our main findings suggest that the intrinsic preference for group members (*GIE*) is the most reliable explanation of social identity. In-group-favoritism cannot instead be accounted for by changes in payoff expectations, as in-group favoritism is neither caused nor mediated by beliefs, as predicted by *BDE* and *BME*, respectively.

The paper is organized as follows. Section 2 briefly discusses the related literature on social identity and on discrimination, with the aim to outline the most relevant contributions for our research. Section 3 introduces the experimental design and the experiment procedures. Section 4 presents and discusses the results. Section 5 concludes.

## 2. Related literature

The pioneering experiments by Tajfel (1970) and Tajfel *et al.* (1971) were the first ones to document that people reward more in-group than out-group members by seeking to maximize either the in-group outcomes, or the differences between in-group and out-group outcomes. These studies showed that this form of intergroup discrimination is basic and that it emerges even under a minimal setting (Minimal Group Paradigm, MGP).[7] Tajfel and coauthors' finding fueled interest in social identity and opened up a new field of research on intergroup bias and discrimination.

In an extensive study based on controlled experiments, Chen and Li (2009) found that in-group members gift, reciprocate, forgive and maximize overall efficiency significantly more than out-group members. Chen and Li (2009) assigned participants to groups either on the basis of their aesthetic preferences for paintings or randomly and did not find statistically significant differences between the behavior of participants assigned according to the two different procedures.[8] Kranton and Sanders (2017) and Kranton *et al*. (2018) compare the outcomes obtained in a minimal group setting to those stemming from a political group aggregation. They found that a large set of subjects show ingroup bias in both. This indicates that bias depends on group division *per se* rather than group identity. They also found robust heterogeneity in behavior in group contexts: some people show ingroup bias, while others do not. In dictator and trust games,

[6] See, among others, Vanberg (2008), Ellingsen *et al*. (2010), Di Bartolomeo *et al*. (2019a, 2019b).

[7] According to the MGP, experiment participants: i) are randomly assigned to groups; ii) are not allowed to communicate; iii) do not know the members of their in-group and of the out-group; iv) have no vested interest in serving their group. Pechar and Kranton (2017) provide a meta-analysis of experiments built on the MGP in the Tajfel *et. al*. (1971) tradition.

[8] Charness *et al*. (2007) did not find any effect of group membership on individual behavior when using the MGP. By contrast, group membership significantly affects individual behavior in other treatments where groups are more salient.

evidence for in-group favoritism was found, among others, by Fershtman and Gneezy (2001), Ben-Ner *et al*. (2009), Whitt and Wilson (2007), Hargreaves *et al*. (2009), Liebe and Tutic (2010), Ioannou *et al*. (2012), and Pan and Houser (2013, 2019).[9]

Closely related to our aim, recent studies explored the role played by beliefs in in-group discrimination. Among them, the experiments designed by Güth *et al.* (2009), Daskalova (2018), and Ockenfels and Werner (2014) are especially worth mentioning. It should be also said that the importance of exploring the role of psychological factors, such as beliefs, has been stressed by many economists and psychologists, including Chen and Li (2009) and Kranton *et al*. (2018). According to Ockenfels and Werner (2014), a fundamental open question is "does social identity change behavior by influencing agent's expectations about fellow ingroup members' behavior by changing the agent's preferences?"

Güth *et al.* (2009) deepened an explanation which is consistent with guilt aversion. As people dislike to let the others' expectations down (Charness and Dufwenberg, 2006), favoritism may emerge if group members believe that the other members believe that group members behave in a more altruistic way with individuals belonging to their own group than with people belonging to other groups. By using dictator and public good games, they provided evidence, based on the observed correlations between dictators' self-reported beliefs and in-group favoritism, which is consistent with guilt aversion. Yet, correlations between transfers and beliefs may be prone to false consensus effects (Ross *et al*., 1977) and provide no conclusive evidence for beliefs to causally affect behavior.

Daskalova (2018) examined individuals' discriminatory behavior in joint decisions and pointed out the potential relevance of beliefs: "expectations of favoritism can be self-fulfilling and can lead to positive discrimination in favor of the own group in joint decisions even if the individual decision makers do not positively discriminate in favor of the own group when deciding alone" (Daskalova, 2018, p. 239). Contrary to our aims, her experiment is not however designed to disentangle the motivations of in-group favoritism.

Finally, by manipulating beliefs, Ockenfels and Werner (2014) showed that in-group favoritism may be partly belief-dependent. Using a dictator game, they found that substantially less in-group favoritism emerges if the dictator is informed that the recipient is unaware of the shared group membership. In one of their experiments (experiment 2), dictators can strategically manipulate the recipients' beliefs. Before the transfer occurs, they may choose to learn whether the opponent shares the dictators' own group identity. Two treatments are then considered. In the first treatment, if the dictator decides to be informed about the recipient's group identity, the same information will be given to the recipient. In the second treatment, the

[9] Other studies are surveyed in Chen and Li (2009, Section I) and Balliet *et al*. (2014).

same information will not be shared with recipients. The bias towards in-group matches turns out to be substantially smaller when the recipient is not informed about the dictator's group affiliation.

Ockenfels and Werner (2014) derive from these findings that in-group favoritism may be partly belief-dependent, because the dictators' second-order beliefs and the recipients' expectations are higher when information is shared. Therefore, whereas in the two treatments dictators have different second-order beliefs, the results could be driven by a self-selection problem due to the fact that the choice of disclosing information can be motivated by different strategic considerations.[10] In order to solve this problem, we elaborate on the exogenous mechanism of belief manipulation proposed by Ockenfels and Werner (2014), to construct the subtler and more articulated device that is described in the next section.

## 3. Experimental design and procedures

### *3.1 Experimental design and hypotheses*

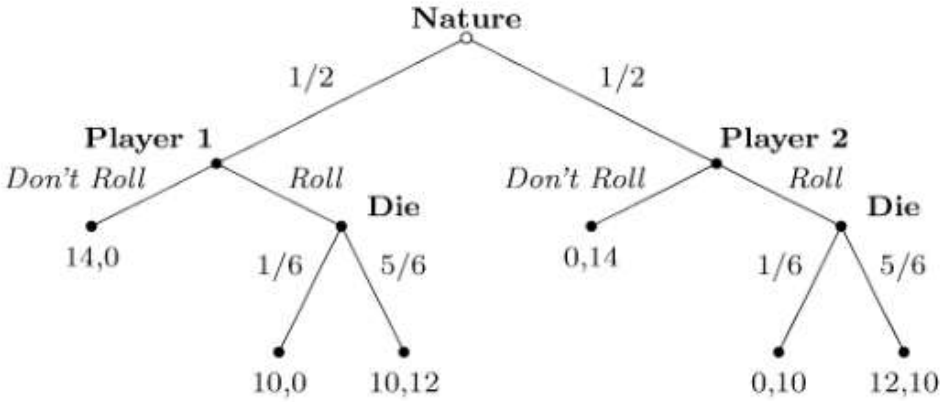
Our experiment is based on the binary-choice random-dictator game described in Figure 1.[11] In this game, *Nature* first randomly determines whether player 1 or 2 will be the *dictator*. The other player will be the *recipient*. The dictator chooses between two actions: *Don't Roll* and *Roll*. Choosing *Don't Roll*, the dictator receives 14 tokens and the recipient receives nothing; choosing *Roll*, the dictator receives 10 tokens, whereas the recipient receives 12 tokens with probability 5/6 and nothings with probability 1/6.[12]

[10] Ockenfels and Werner (2014) show that dictators strategically use information disclosure, i.e., they behave differently in the two treatments.

[11] We use a dictator game as in Guth *et al*. (2009) and Ockenfels and Werner (2014).

[12] The payoffs and the names of choices in subgames match those used in the well-known Charness and Dufwenberg's (2006) trust game. Our setup also guarantees that recipients cannot observe an egoistic action, i.e., *Don't Roll*, by the sender. As this experiment is part of a wider project carried out at the *CIMEO Experimental Economics Lab* of the Sapienza University of Rome on effort determinants in asymmetric information contexts, in future research we plan to compare its results with those stemming from designs based on different mechanisms that could provide incentives for altruistic behavior (e.g., repeated interaction, promise giving, non-binding agreements, etc.).

**Figure 1. A binary-choice random-dictator game**




If all the players are assigned to one of two different groups in a random way before the game starts, then the expected outcome, which was actually obtained in our experiments,[13] will be that the subjects' behavior is affected by group membership and that favoritism occurs. A dictator is more likely to *Roll* when he/she is paired with a recipient belonging to the same group (good match).

Evidence for in-group favoritism is also related to beliefs. The increase in the rate at which dictators in good matches choose *Roll* is associated to an increase in both recipients' beliefs that they will do so (first-order beliefs) and dictators' beliefs about the recipients' beliefs that they will do so (second-order beliefs). This correlation between *Roll* rates and second-order beliefs is not however sufficient to fully understand the rationale of in-group favoritism, as causations can run in both directions. To this goal, it is also necessary to investigate:

1. if the increased frequency of *Roll* choices in good matches is driven by changes in second-order beliefs eventually induced by group membership (*BDE*);
2. in the absence of evidence in favor of *BDE*, whether a preference for group membership is mediated or not by beliefs (*BME* vs. *GIE*).

This investigation requires to introduce independent (exogenous) variations in both group membership (to test the first issue) and second-order beliefs (to test the second issue). These variations are here obtained by modifying the random-dictator game by employing the switching mechanism introduced by Vanberg (2008).[14] This modification is common knowledge for all participants.

[13] The results are reported in Section 4.1 and Appendix A.

[14] Following Vanberg (2008) and Ellingsen *et al*. (2010), Di Bartolomeo *et al*. (2019a, 2019b) introduced similar mechanisms to evaluate the causal nexus linking second-order beliefs to promise keeping.

In our experiment, all participants are first assigned to the "Reds," or to the "Blues."[15] Pairs of subjects are then formed randomly, Nature chooses the paired-subject roles and all dictators and recipients observe the color of their partners. Next, some of the recipients are randomly re-matched with a new dictator, according to a given switching probability, which is known from the beginning. After the re-matching, each dictator (be she involved or not in a switch) chooses between *Roll* and *Don't Roll*, and the payoffs are determined according to Figure 1.

The exogenous variation in group membership is obtained by introducing asymmetric information between dictators and recipients for a given switching probability.

1. After the re-matching, all recipients know that some of them could have been switched, but they do not observe whether the switch actually occurred. By contrast, dictators know if they were switched and, in such a case, they are also informed about the color of the new recipient and whether she was initially in a good or a bad match.
2. The above information design leads to an exogenous variation in group membership since it implies that recipients' first-order beliefs and dictators' second-order beliefs are independent of whether they were effectively switched or not. Specifically, the recipient's first-order beliefs are affected by the initial match (good or bad), but they are independent of the switch/no-switch condition, which cannot be observed by recipients. As dictators know that recipients do not observe the switch, it follows that their second-order beliefs do not depend on the switch too.

The above exogenous variations in group membership allow us to test *BDE*. We focus on the subsets of switched dictators playing with new recipients who were initially in the same situation (given the switching probability, they were all either in a good or in a bad match). These switched dictators could play in both good or bad matches, but they have the same second-order beliefs, which only depend on the given switching probability and the given kind (good or bad) of recipients' initial match. *BDE* predicts that behavior is only driven by beliefs; thus, in each subset, it implies no difference in the generosity of switched dictators in good and bad matches, as there they have the same second-order beliefs.

Formally, the test for *BDE* can be stated as follows.

**H1:** *Testing the BDE*. Taking as given the switching probability and the kind (good or bad) of recipients' initial match (which affect beliefs), we compare the behavior of switched dictators who are re-matched with out-group recipients (bad match) with that of switched dictators who are re-matched with in-group recipients (good match). If the dictators' behavior is not the same, then the *BDE* cannot hold.

[15] See the next subsection for details.

If the results of the experiments do not confirm H1, the observed in-group bias would not be consistent with *BDE*. This does not however fully rule out beliefs and we must further investigate whether group identity is *mediated*, or not, by beliefs, i.e., we need to test *BME* vs. *GIE*. To this aim, we no longer need to produce an exogenous variation in group membership, but rather an exogenous variation in expectations, to compare the behavior of dictators in good matches who hold different second-order beliefs. We produce this exogenous variation in expectations by using the extension of Vanberg's (2008) methodology introduced by Di Bartolomeo *et al*. (2019a) in the context of promise keeping. This extension is based on the introduction of two different, given and known, switching probabilities (high and low). If group identity is consistent with in-group favoritism, then the recipients' first-order and the dictators' second-order beliefs will depend on the value of the switching probability.

Assuming that *BDE* was ruled out by H1, then, for the same beliefs, we know that dictators are more generous only when they are in good matches. Focusing on the recipients in good matches, those facing the lower switching probability have higher beliefs compared to the others. Second-order beliefs follow the first-order ones because of the employed information structure: focusing on good matches, if the switching probability is low (high) enough, second-order beliefs are high (low).[16] This explains why our treatment variable, the manipulation of switching probabilities, produces exogenous variations in (first- and) second-order beliefs.

We use the exogenous variation in beliefs to test *BME* vs. *GIE*:

i. a higher *Roll* rate among non-switched dictators with higher second-order expectations would imply that the dictators' behavior is causally affected by their expectations, as predicted by *BME*;
ii. the same *Roll* rate among non-switched dictators, regardless of the second-order expectations, would imply that the dictators' behavior is not causally affected by their expectations, as predicted by *GIE*.

More synthetically, if H1 does not hold, the comparison between *BME* and *GIE* requires to check whether group identity is mediated by beliefs, as suggested by *BME*, or it is not mediated by beliefs, as suggested by *GIE*. Formally:

[16] Of course, marginal differences in the probabilities lead to marginal difference in beliefs. By contrast, looking at the limiting cases (1 and 0) for the two switching probabilities, the largest difference in beliefs is obtained. Limiting cases cannot however be used because, in such a case, recipients can easily infer if they were switched or not.

**H2:** *Testing the BME vs. GIE* (two tests).

1. We consider non-switched dictators in good matches and compare their behavior under a low switching probability (high second-order beliefs) with that under a high switching probability (low second-order beliefs).
2. We consider switched dictators in good matches who were also initially in good matches and compare their behavior under a low switching probability (high second-order beliefs) with that under a high switching probability (low second-order beliefs).

In each of the above cases, if the dictators' behavior is the same, then beliefs will not matter, *BME* will not hold and *GIE* will be the explanation with the greatest explanatory power.

### *3.3 Procedures*

The experiment was conducted in November 2017 at the *CIMEO Experimental Economics Lab* of the Sapienza University of Rome.[17] The design involved 384 students of this University (12 sessions, 10 rounds each, 32 subjects each), recruited using an online system. Upon arrival, subjects were randomly assigned to 32 isolated computer terminals.[18] Three assistants handed out instructions[19] and checked that participants correctly followed the procedures. Before playing any game, subjects filled out a short questionnaire testing their comprehension of the experiment.

The experiment consisted of three stages: group assignment (pre-session); experimental session; final payment (post session). At the beginning of the game, subjects were randomly assigned to one of two groups, Red and Blue, and they remained in the same group until the end of the experiment (i.e., groups did not change throughout the experiment). All subjects were informed about their group membership and about the membership of the other participants.

After the group assignment, subjects played the experimental session, which consisted of ten rounds, with perfect stranger matching. Each round implemented the following sequence of five stages:

1. *Role assignment.* At the beginning of the round, roles (dictator or recipient) were randomly assigned for each pair and subjects were informed of that. All dictators observed the *color* of the paired recipient and vice versa.
2. *Revelation of the switching probability*. The switching probability was announced (it could be either 12.5% or 87.5%).

[17] This is the acronym for *Center for Interpretation and Modeling of Experimental Observations.*

[18] The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

[19] See the online Supplementary Material.

3. *Belief elicitation*. This stage has two parts:
    a. First-order beliefs: each recipient was asked to guess about his/her payoff.
    b. Second-order beliefs: each dictator was asked to guess the guess of the recipient with whom he/she had formed a pair.
4. *Switching*. Some pairs were switched, accordingly to the round-given-switching probability. Only dictators were informed of whether a switch occurred (recipients were not informed off whether they had been switched). Dictators with switched recipients were also informed of: 1) the group's color of the new recipients; 2) the group's color of the previous partners (dictators) of their new recipients.
5. *Dictators' action*. All dictators made their choice: *Roll* or *Don't Roll*. All subjects were informed of their payoff for the round. Recipients could not infer the dictator's choice when their payoffs were zero.[20]

Finally, at the end of the session, subjects were paid. All subjects received a fixed show-up fee of 2.50 tokens. One of the rounds was randomly chosen for payment. The payoffs shown in Figure 1 were computed in tokens (where 1 token = 0.5 euro). At the end of each session, incentives for beliefs elicitation were provided for all rounds, except the one chosen for payment, implying that subjects had no incentive to hedge against bad outcomes and thus to misreport their beliefs.[21]

It is worth stressing that, in line with Charness and Dufwenberg's (2006), we elicit beliefs before dictators make their choice, that is, before they know whether they have been switched or not. Exogenous variations in their expectations can hence be due only to a change in the switching probability, given the initial match. We assume that a change in the switching probability produces also exogenous variations in switched dictators' expectations.[22] This is reasonable because dictators know that recipients do not observe whether a switch occurred and so their first-order beliefs should depend on the switching probability and on the initial match, but not on the switch condition.[23]

## 4. Results

[20] Recipients could obtain a zero payoff in two cases: (i) their dictator had chosen *Don't Roll*; (ii) their dictator had chosen *Roll* and the outcome of the die-roll was "1."

[21] Appendix B describes in detail how beliefs were elicited and computed.

[22] Recall that we elicit the beliefs of all dictators before the switch takes place, whereas we do not elicit the beliefs of *switched* dictators regarding the beliefs of their *re-matched* recipients.

[23] See Di Bartolomeo *et al*. (2019a) for a more detailed discussion on the elicitation timing in the procedure here employed.

### *4.1 Matching, beliefs and Roll rates*

Table 1 reports the averages of dictators' second-order beliefs and *Roll* rates, the standard deviations and the number of observations. The rows of the table also distinguish: (i) dictators who were matched with an out-group member (row (A)) from (ii) dictators who were matched with an in-group member (row (B)). Tables also show standard deviations (s.d.) and observations (obs.).

The table provides evidence in favor of in-group favoritism.[24] The average *Roll* rate of in-group members is higher (50%) than that of out-group members (34%) [i.e., $Z$=2.98, $p$=0.003]. The averages comparisons are always performed by using non-parametric tests. All the statistics reported adopt the Wilcoxon signed-rank, which compares averages at the session level. Since our data are independent at the session level, but not at the individual level, the test accounts for this data structure.[25]

The figures contained in the cells also show that the average second-order beliefs of in-group members (53%) are higher than those of out-group members (40%) [i.e., $Z$=3.06, $p$=0.002].[26] We hence confirm the existence of a correlation between behavior and second-order beliefs (Guth *et. al.*, 2009; Ockenfeld and Werner, 2014) but, in order to identify the direction of causality, we must look at exogenous variations of group membership.

**Table 1. Matching, beliefs, and *Roll* rates (1920 obs.)**

| | DICTATORS | |
|---|---|---|
| | AVERAGE ROLL RATES | SECOND-ORDER BELIEFS |
| MATCHING GROUPS | (1) | (2) |
| (A) OUT GROUP | 0.34<br>(s.d. 0.48, obs. 912) | 0.40<br>(s.d. 0.32, obs. 912) |
| (B) IN GROUP | 0.50<br>(s.d. 0.50, obs. 1008) | 0.53<br>(s.d. 0.33, obs. 1008) |

### *4.2 Expectations and switching probabilities*

Before exploring the causal effect of group membership and beliefs, we must establish that our design is indeed able to induce exogenous variations in first- and in second-order beliefs. Table 2 and Table 3 present

[24] Note that our switching mechanism tends to be unfavorable for our tests, as Ockenfels and Werner (2014) documented that there is substantially less in-group favoritism if the dictator is informed that the recipient is unaware of the shared group membership.

[25] All outcomes obtained from the reported tests are also checked by using the Fligner-Policello test for small sample, which confirms them in all the cases (results are available upon request).

[26] It is worth noting that we would obtain the same results if we looked at the first-round data only (see Appendix C: Table 3C).

(first-order) recipients' and (second-order) dictators' average beliefs about dictator's decision to choose *Roll*. In the tables, columns refer to the cases of a low (column 1) and a high (column 2) switching probability; rows refer to matching: in-group refers to all the pairs where the dictator and the recipients were initially assigned to the same group color (row B); out-group refers to the other pairs (row A).

Consider recipients first. The comparison of first-order beliefs by row (A and B respectively), taking into account low and high switching probabilities, confirms the existence of an exogenous variation in first-order beliefs. Specifically, the recipients who are matched with an out-group member under a low switching probability to change partner expect that 36% of dictators will chose to roll, whereas the expected fraction increases to 45% among the recipients who are matched with an out-group member under a high switching probability (i.e., 36% vs. 45%: $Z$=2.66, $p$=0.007).

Similarly, the first-order beliefs of recipients who are matched with an in-group member under a low switching probability to change partner assign the value 58% to the fraction of dictators who will chose to roll, whereas the recipients who are matched with an in-group member under a high switching probability expect the fraction of rolling dictators to be equal to 47% (i.e., 58% vs. 47%: $Z$=3.06, $p$=0.002).

It is also worth noting that under a low probability of changing partner the first-order beliefs of recipients who are matched with an in-group member are characterized by a greater average than that of recipients who are matched with an out-group member (i.e., 58% vs. 36%: $Z$=3.06, $p$=0.002), whereas under a high switching probability the first-order beliefs of the two types of recipients are not statistically different from each other (i.e., 47% vs. 45%: $Z$=1.26, $p$=0.209).

**Table 2. Recipients' first-order beliefs (1920 obs.)** *

| | SWITCHING PROBABILITY | |
|---|---|---|
| | LOW (12.5%) | HIGH (87.5%) |
| MATCHING GROUPS | (1) | (2) |
| (A) OUT GROUP (I.E., BAD MATCHES) | 0.36 (s.d. 0.32, obs. 384) | 0.45 (s.d. 0.31, obs. 528) |
| (B) IN GROUP (I.E., GOOD MATCHES) | 0.58 (s.d. 0.33, obs. 576) | 0.47 (s.d. 0.31, obs. 432) |

* The first-order beliefs are elicited for all recipients.

Similarly, by looking in Table 3 at second-order beliefs and comparing rows A and B, under a low (column 1) and a high (column 2) switching probability, we can confirm the existence of an exogenous variation in expectations. Specifically, the dictators who are matched with an out-group member under a low probability of changing partner believe that the recipients expect dictators will *Roll* with a probability equal to 35%, whereas the same expectation increases to 43% among the dictators who are matched with

an out-group member under a high switching probability (i.e., 35% vs. 43%: $Z$=2.74, $p$=0.006). The second-order beliefs of dictators who are matched with an in-group member under a low switching probability to change partner are instead characterized by an expected probability to *Roll* equal to 59%, whereas this expectation is 46% among the dictators who are matched with an in-group member under a high switching probability (i.e., 59% vs. 46%: $Z$=2.98, $p$=0.003).

Also note that under a low probability of changing partner, the second-order beliefs of dictators who are matched with an in-group member are characterized by a greater expectation on *Roll* than that of dictators who are matched with an out-group member (i.e., 59% vs. 35%: $Z$=3.06, $p$=0.002). The same may seem to occur under a high switching probability, but in this case the difference is not statistically significant (i.e., 46% vs. 43%: $Z$=0.78, $p$=0.432).

**Table 3. Dictators' second-order beliefs (1920 obs.)** *

| | SWITCHING PROBABILITY | |
|---|---|---|
| | LOW (12.5%) | HIGH (87.5%) |
| | (1) | (2) |
| MATCHING GROUPS | | |
| (A) OUT GROUP | 0.35<br>(s.d. 0.32, obs. 384) | 0.43<br>(s.d. 0.32, obs. 528) |
| (B) IN GROUP | 0.59<br>(s.d. 0.34, obs. 576) | 0.46<br>(s.d. 0.31, obs. 432) |

* Second-order beliefs are elicited for all dictators before they know whether they have been re-matched or not.

### *4.3 Are beliefs the determinant of in-group favoritism?*

We begin by investigating whether our data provide some evidence for *BDE*. As explained above, in order to test H1, we have to focus on switched dictators. The information needed to test *BDE* is reported in Table 4, which contains the average *Roll* rates of switched dictators, together with the standard deviations and the number of observations. The eight cells/categories of the table can be considered, by row, across two dimensions: dictators who are re-matched with (i) out-group (row A) and (ii) in-group recipients (row B). These are matched by columns with other two dimensions related to the new recipients' history (initial match of recipient's), distinguishing those who were initially matched with (i) out-group and (ii) in-group members, under high and low switching probability (columns 1-2 and 3-4), respectively.

**Table 4 – Exogenous variation in group membership - Switched dictators' *Roll* rates (960 obs.).**

| | HIGH SWITCH PROBABILITY | LOW SWITCH PROBABILITY |
|---|---|---|

| | (840 OBS.) | | (120 OBS.) | |
|---|---|---|---|---|
| | INITIAL MATCH (RECIPIENT) | | INITIAL MATCH (RECIPIENT) | |
| | OUT GROUP | IN GROUP | OUT GROUP | IN GROUP |
| FINAL MATCH | (1) | (2) | (3) | (4) |
| (A) OUT GROUP | 0.27 (s.d. 0.44, obs. 264) | 0.27 (s.d. 0.44, obs. 192) | 0.25 (s.d. 0.45, obs. 12) | 0.28 (s.d. 0.45, obs. 60) |
| (B) IN GROUP | 0.54 (s.d. 0.50, obs. 216) | 0.54 (s.d. 0.50, obs. 168) | 0.39 (s.d. 0.49, obs. 36) | 0.75 (s.d. 0.45, obs. 12) |

The behavior of *switched* dictators is examined when an exogenous variation of group membership is produced. According to H1, our test proceeds as follows. For any given switching probability and recipients' initial match (i.e., column),[27] we compare the behavior of out-group dictators (bad re-matched) to that of in-group dictators (good re-matched). Formally, H1 is tested by comparing the average roll rates of row A to that of row B for any column, obtaining then four tests. It must be noted that the most accurate tests are those based on the case of high switching probability (the two tests sub 1. below) since they are characterized by a much higher number of observations.[28]

The results of the four tests are as follows.

1. Under a *high switching probability*, dictators are more likely to *Roll* when they are re-matched with a member of their group. This occurs both when their new recipients were initially in bad matches (column 1) and when they were in good matches (column 2) [i.e., respectively, 27% vs. 54%: $Z$=2.98, $p$=0.003 (test 1.1); and 27% vs. 54%: $Z$=2.98, $p$=0.003 (test 1.2)].
2. Under a *low switching probability*, we obtain two different outcomes. Differently from sub 1., dictators are not more likely to *Roll* when they are re-matched with in-group members who were initially matched with out-group dictators (column 3) [i.e., 25% vs. 39%: $Z$=1.44, $p$=0.149 (test 2.1)]. By contrast, as in sub 1., dictators are more likely to *Roll* they are re-matched with in-group members who were initially matched with in-group dictators (column 4) [i.e., 28% vs. 75%: $Z$=2.62, $p$=0.009 (test 2.2)].

The (most accurate) tests sub. 1 reject H1 since they provide evidence that in-group favoritism cannot be explained only by beliefs and so they argue against *BDE*. The same is found in 2.2. By contrast, 2.1 seems to suggest that group identity does not always play a role, but the very small number of observations

[27] Recall that the switch is not observed by the recipients and so dictators' second-order beliefs only depend on the observed initial match of their recipients (reported in the columns).

[28] This occurs by construction of our design. We are focusing on switched dictator to test BDE. By restricting to these players the data on Table 4, it follows that the tests based on the high switching probability are characterized by a much higher number of observations (840 on the overall) than those based on the low switching probability (120).

which characterizes this case leads us to the conclusion that the dictators' behavior does not generally appear to be causally affected by expectations.[29]

Summing up, Table 4 provides evidence that in-group favoritism cannot be explained only by beliefs and hence runs against *BDE* (H1 is rejected). Beliefs might however mediate the dictators' behavior within in-groups, as predicted by *BME*. This possibility is explored in the next section.

### *4.4 Is in-group favoritism mediated by beliefs?*

We now test whether in-group favoritism is mediated by beliefs or not (H2), that is, whether our evidence is in favor of *BME* or of *GIE*. To carry out this test, we need to use the exogenous variation in second-order beliefs to compare, *caeteris paribus*, the behavior of in-group dictators holding different second-order beliefs. The exogenous variation is obtained by considering the in-group dictators matched with recipients with different initial matching (histories) and playing with different switching probabilities, that affect the dictators' second-order beliefs.

As explained in Section 3.1, H2 is tested by focusing on the in-group dictators' behavior reported in Table 5. This table reports their average *Roll* rates and its cells/categories are structured along two dimensions: rows distinguish between non-switched (A) and switched dictators (B); columns are related to the initial matchings of their recipients (good or bad matches) by switching probabilities (high or low), which in turns are the determinants of their second-order beliefs. Looking at row A, recall that, as shown in Table 3, row B, the second-order beliefs of non-switched dictators under high (column 2) and low (column 4) switching probability are statistically different.

Consider first the non-switched dictators in good matches in Table 5, i.e., row A, column 2 vs. column 4. The different second-order beliefs notwithstanding, we get that the average *Roll* rate of dictators under a low switching probability is not statistically different from the same average rates under a high switching probability (i.e., 61% vs. 58%: $Z$=0.31, $p$=0.753).[30] This shows that the behavior of in-group dictators does not seem to be different when their second-order expectations are different. In other words, the non-switched dictators' behavior does not appear to be causally affected by their expectations.

[29] This is the case with the smallest number of observations (12 vs. 36 distributed in 12 sessions). Remember that we perform non-parametric tests at the session levels.

[30] It is worth noting that a similar result holds for non-switched out-group dictators (not reported in Table 5): the average *Roll* rate of dictators under a low switching probability is not statistically different from that of dictators under a high switching probability (i.e., 30% vs. 38%: $Z$=0.47, $p$=0.637).

**Table 5 – Exogenous variation in beliefs - In-group dictators' *Roll* rates (1008 obs.).**

| | HIGH SWITCH PROBABILITY | | LOW SWITCH PROBABILITY | |
|---|---|---|---|---|
| | INITIAL MATCH (RECIPIENT) | | INITIAL MATCH (RECIPIENT) | |
| | OUT GROUP | IN GROUP | OUT GROUP | IN GROUP |
| SWITCHING CONDITION | (1) | (2) | (3) | (4) |
| (A) NON-SWITCHED | | 0.58 (s.d. 0.50, obs. 72) | | 0.61 (s.d. 0.49, obs. 504) |
| (B) SWITCHED | 0.54 (s.d. 0.50, obs. 216) | 0.54 (s.d. 0.50, obs. 168) | 0.39 (s.d. 0.49, obs. 36) | 0.75 (s.d. 0.45, obs. 12) |

Similar tests were carried out for row B and the same results obtained. First, focusing on switched dictators in good matches who are re-matched with a recipient who initially was in a good match (row B, columns 2 vs. 4), we observe that their average *Roll* rate under a low switching probability is not statistically different from that under a high switching probability (i.e., 54% vs. 75%: $Z$=1.40, $p$=0.162). Second, the results of a further test focusing on the switched dictators playing under a low switching probability (row B, columns 3 vs. 4)[31] confirm that the switched dictators' behavior is not causally affected by their expectations.[32]

Our conclusion is that the results from Table 5 suggest that in group membership matters *per se*. As all the tests reject H2, we do not find evidence in favor of *BME* and cannot hence confirm the results obtained by Ockenfels e Werner (2014), who find evidence supporting the view that in-group favoritism is partially belief-dependent. Our explanation of these different findings is that those by Ockenfels e Werner (2014) were based on a variation in expectations which was not exogenous, but based on strategic choices made by dictators in different contexts, whereas ours are based on an exogenous variation in expectations. These

[31] Note that the second-order beliefs of in-group and out-group dictators under low switching probability (columns 3 and 4) are statistically different. See Table 3, column 1.

[32] Although based on few observations, the same is observed by comparing the average *Roll* rates of dictators who are re-matched either with initially in-group (0.75) or initially out-group (0.39) recipients [75% vs. 39%: $Z$=1.90, $p$=0.058]. Further evidence in favor of the fact that beliefs do not matter for in-group favoritism can be drawn from the outcomes of Table 5. Out of the tests designed for H2, it should be noted that, given the switching probability, the average roll rate of non-switched dictators is not statistically different from that of switched dictators (playing with a recipient initially badly matched) although second order-beliefs are not the same. Under a high switching probability, the test is: 58% vs. 54%: $Z$=0.71, $p$=0.480; under a low switching probability the test is: 61% vs. 39%: $Z$=1.88, $p$=0.060. Similarly, switched dictators (playing with a recipient initially badly matched) also have different second-order beliefs under a high and a low switching probability, but their average roll rates are not statistically different (54% vs. 39%: $Z$=1.30, $p$=0.195).

leads us to believe that Ockenfels e Werner's (2014) results are likely to be driven more by a self-selection bias related to the strategic use of information disclosure than by beliefs.

### *4.5 Investigating individual data*

We use individual-level panel data to further test the different explanations of group favoritism previously addressed. In Table 5, we estimate a random intercept probit model using GLLAMM (Stata). The probit panel regression is based on 1920 observations. The dependent variable is the probability that the dictator chooses *Roll*. Standard errors are indicated in the last column. By one (two) asterisk, we indicate significance at 1% (5%) level.

The estimation considers the gender of the dictator and whether the dictator was switched. It also accounts for the matching: "In Group" means that the dictators and the recipients belong to the same group. "Recipient initial in group" is equal to one if the recipient was matched with a partner of the same group before the switch occurred. According to our design, the interaction between "Low switching probability" and "Recipient's initial in group" captures high second-order beliefs against lower beliefs. Finally, "Round" accounts for the experiment dynamics.

**Table 6 – Estimate of panel regression**

| | Coef. | Standard error |
|---|---|---|
| Male | -0.007 | 0.14 |
| Switch | -0.204 | 0.15 |
| In Group | 0.646* | 0.13 |
| Recipient's initial in group | 0.103 | 0.13 |
| In Group × Switch | 0.176 | 0.14 |
| Low switching probability × Recipient's initial in group | -0.143 | 0.17 |
| Low switching probability × Recipient's initial in group × In Group | 0.300 | 0.19 |
| Round | 0.005 | 0.01 |
| Constant | -0.61* | 0.12 |
| | | |
| Log likelihood | -1176.77 | |

Estimation results are in line with the conclusions derived in the main text. Specifically, dictators who were matched with recipients belonging to the same group were significantly more likely to *Roll* and this effect was observed independently of whether their partners were switched or not. Switching did not have a significant impact on behavior both *per se* and on the dictators matched with recipients belonging to the same group. This is not surprising, as our results suggest that group identity explains in-group favoritism *per se*.

Once we control for low switching probability and recipient's initial group (which are associated to high expectations), we find that expectations do not matter. This suggest that, as argued in the main text, social identity is not mediated by expectations. Expectations matter neither in general ("Low switching probability × Recipient's initial in group" is not significant) nor within the dictators who are matched with recipients belonging to the same group ("Low switching probability × Recipient's initial in group × In Group" is not significant).

## 5. Conclusions

In this paper we put to a test three possible explanations of in-group favoritism: a) People have an intrinsic preference for members of the own group (*Group Identity Explanation, GIE*); b) people dislike to let the others' expectations down and in-group favoritism emerges as the equilibrium of a psychological game where social identity shapes beliefs (*Beliefs Driven Explanation, BDE*); c) people care only about the beliefs of their own group members and in-group favoritism is the result of a sort of intersection of the two previous explanations (*Belief-Mediated group identity Explanation, BME*).

The three explanations differently emphasize the role played by second-order beliefs and intrinsic preferences for people belonging to their own group. All of them imply a positive correlation between second-order beliefs and in-group favoritism, but are driven by different causality. Testing them is thus difficult, even if second-order beliefs are elicited. In order to overcome this difficulty, we proposed a design that disentangles the effects produced by second-order beliefs from those produced by group membership through a double exogenous variation in both beliefs and group membership. This design allowed us to test: (i) if beliefs can explain in-group favoritism independently of group membership (the *BDE* holds); (ii) if this were not the case, whether in-group bias is independent of expectations (the *GIE* holds), or is mediated by beliefs (the *BME* holds).

The experiment we carried out provided the following results. First, in-group-favoritism is not accounted for by changes in payoff expectations. This suggests that in-group favoritism is not "caused" by beliefs and the *BDE* cannot be supported. Second, in-group favoritism is not mediated by second-order beliefs and the *BME* cannot be supported. Third, in line with the formalizations proposed by, e.g., Akerlof and Kranton (2000), Chen and Li (2009), and Chen and Chen (2011), our experiment supports an explanation based on intrinsic preferences for group members and suggests to single out the *GIE* as the most powerful explanation of in-group-favoritism.

Our results consequently suggest that policy interventions to fight discrimination should aim to change, not the expectations of the ones who are discriminated, but the preference for group identity of those who discriminate. When the effects of group identity are positive for society, policy measures should instead

aim at strengthening such preference. The policy fields potentially interested by our findings hence span, on the one side, e.g., from bullying in school to discrimination against immigrants and, on the other side, e.g., from team work to partnership in research activity.

Our conclusions raise further research questions, as we are now interested in understanding whether beliefs may re-enter the scene when dynamics is considered, for example when repeated interactions among members within groups,[33] or loyalty to the group are taken into account. We are also interested in verifying the robustness of our results in other situations characterized by more complex interactions than those considered here. Studies in other contexts show in fact that the effects of beliefs on choices depend on the specific context and on the interpersonal dynamics considered. These represent for us fruitful directions for future research.

## References


Akerlof, G. A. and R. E. Kranton (2000), "Economics and identity," *The Quarterly Journal of Economics* 115: 715–753.

Attanasi, G. and R. Nagel (2008), "A survey of psychological games: Theoretical findings and experimental evidence," in A. Innocenti and P. Sbriglia (eds.), *Games, Rationality and Behavior. Essays on Behavioral Game Theory and Experiments*, Palgrave McMillan: 204–232.

Attanasi, G., P. Battigalli and E. Manzoni (2016), "Incomplete-information models of guilt aversion in the trust game," *Management Science* 62: 648–667.

Babcock, P., Bedard, K., Charness, G., Hartman, J. and H. Royer (2015), "Letting down the team? Social effects of team incentives," *Journal of the European Economic Association*, 13: 841–870.

Bacharach, M., G. Guerra and D. J. Zizzo (2007), "The self-fulfilling property of trust: An experimental study," *Theory and Decision*, 63: 349–388.

Battigalli, P. and M. Dufwenberg (2007), "Guilt in games," *American Economic Review*, 7: 170–176.

Battigalli, P. and M. Dufwenberg (2009), "Dynamic psychological games," *Journal of Economic Theory*, 141: 1–35.

Balliet, D., Y. Wu and C. K. W. De Dreu (2014), "In-group favoritism and cooperation: A meta-analysis," *Psychological Bulletin*, 140: 1556–1581.

Bellemare C., A. Sebald and M. Strobel (2011), "Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models," *Journal of Applied Econometrics*, 26(3): 437–453.


[33] On the role played by repeated interactions in this context, see, e.g., Balliet *et al.* (2014) and Dorrough *et al*. (2015).

Ben-Ner, A., B. P. McCall, M. Stephane and H. Wang (2009), "Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence," *Journal of Economic Behavior & Organization*, 72: 153–170.

Bracht, J. and T. Regner (2013), "Moral emotions and partnership," *Journal of Economic Psychology*, 39: 313–326.

Braver, S. (1995), "Social contracts and the provision of public goods," in *Social dilemmas: Perspectives on individuals and groups*, ed. by D. Schroeder. New York: Praeger.

Charness, G. and M. Dufwenberg (2006), "Promises and partnership," *Econometrica*, 74: 1579–1601.

Charness, G., L. Rigotti and A. Rustichini (2007), "Individual behavior and group membership," *American Economic Review* 97: 1340–1352.

Charness, G. and U. Gneezy (2009) "Incentives to exercise," *Econometrica*, 77: 909–931.

Charness, G. and A. Rustichini (2011), "Gender differences in cooperation with group membership," *Games and Economic Behavior*, 72: 77–85.

Charness, G., R. Cobo-Reyes and N. Jiméne (2014), "Identities, selection, and contributions in a public-goods game," *Games and Economic Behavior* 87: 322–338.

Chen, R. and Y. Chen (2011), "The potential of social identity for equilibrium selection," *American Economic Review* 101: 2562–2589.

Chen, Y. and S. X. Li (2009), "Group identity and social preferences," *American Economic Review*, 99: 431–457.

Daskalova, V. (2018), "Discrimination, social identity, and coordination: An experiment," *Games and Economic Behavior*, 107: 238–252.

Di Bartolomeo, G., M. Dufwenberg, S. Papa and F. Passarelli (2019a), "Promise, expectations and causation," *Games and Economic Behavior*, 113: 137–146.

Di Bartolomeo, G., M. Dufwenberg, and S. Papa (2019b), "The sound of silence: A license to be selfish," *Economics Letters*, 182: 68-70.

Dorrough, A. R., Glöckner, A., Hellmann, D.M. and I. Ebert (2015), "The development of in-group favoritism in repeated social dilemmas," *Frontiers in Psychology*, 6: 1–12.

Dufwenberg, M. and U. Gneezy (2000), "Measuring beliefs in an experimental lost wallet game," *Games and Economic Behavior*, 30: 163–182.

Ederer, F. and A. Stremitzer (2017), "Promises and expectations," *Games and Economic Behavior*, 106: 161–178.

Ellingsen, T. and M. Johannesson (2004), "Promises, threats and fairness," *The Economic Journal*, 114: 397–420.

Ellingsen, T., M. Johannesson, S. Tjøtta and G. Torsvik (2010), "Testing guilt aversion," *Games and Economic Behavior*, 68: 95–107.

Eckel, C. C. and P. J. Grossman (2005), "Managing diversity by creating team identity," *Journal of Economic Behavior & Organization*, 58: 371–392.

Everett, J. A. C., N. S. Faber and M. Crockett (2015), "Preferences and beliefs in in-group favoritism," *Frontiers in behavioral neuroscience*, 9.

Fershtman, C., U. Gneezy and F. Verboven (2005), "Discrimination and nepotism: The efficiency of the anonymity rule," *Journal of Legal Studies*, 34, 371–396.

Fischbacher, U. (2007), "z–Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10: 171–178.

Goette, L., D. Huffman, S. Meier and M. Sutter (2012), "Competition between organizational groups: Its impact on altruistic and antisocial motives," *Management Science*, 58: 948–960.

Guala, F., L. Mittone and M. Ploner (2013), "Group membership, team preferences, and expectations," *Journal of Economic Behavior & Organization* 86: 183–190.

Guerra, G. and D. J. Zizzo (2004), "Trust responsiveness and beliefs," *Journal of Economic Behavior & Organization*, 55: 25–30.

Güth, W., M. Ploner and T. Regner, (2009), "Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?" *Journal of Economic Psychology*, 30: 814–827.

Hargreaves Heap, S. P. and D. J. Zizzo (2009). "The value of groups," *American Economic Review*, 99: 295–323.

Huang, P. H. and H. M. Wu (1994), "More order without more law: A theory of social norms and organizational cultures," *Journal of Law, Economics, and Organization*, 10: 390–406.

Ioannou, C. A., S. Qi and A. Rustichini (2015), "Group payoffs as public signals," *Journal of Economic Psychology*, 48: 89–105.

Kranton, R. and S. Sanders (2017), "Groupy vs. non groupy social preferences: Personality, region, and political party," *American Economic Review Papers and Proceedings*, 107: 65–69.

Kranton, R., M. Pease S. Sanders, and S. Huettel (2018), "Groupy vs. non-groupy behavior: Deconstructing group bias," Duke University, mimeo.

Khalmetski, K., A. Ockenfels and P. Werner (2015), "Surprising gifts: Theory and laboratory evidence," *Journal of Economic Theory*, 159: 163–208.

Klor, E. F. and M. Shayo (2010), "Social identity and preferences over redistribution," *Journal of Public Economics*, 94: 269–278

Liebe, U. and A. Tutic (2010), "Status groups and altruistic behaviour in dictator games," *Rationality and Society*, 22: 353–380.

Morita, H. and M. Servátka (2013), "Group identity and relation-specific investment: an experimental investigation," *European Economic Review*, 58: 95–109.

Ockenfeld, A. and P. Werner (2014), "Beliefs and in-group favoritism," *Journal of Economic Behavior & Organization*, 108: 453–462.

Ostrom, E., J. Walker and R. Gardner (1992), "Covenants with and without a sword: Self–governance is possible," *The American Political Science Review*, 86: 404–417.

Pan, X. S. and D. Houser (2013), "Cooperation during culture group formation promotes trust towards members of out-groups," *Proceedings of the Royal Society* B: Biological Sciences 280, no. 1762.

Pan, X. S. and D. Houser (2019), "Why trust out-groups? The role of punishment under uncertainty," *Journal of Economic Behavior & Organization*, 158: 236–254.

Pechar, E. and R. Kranton (2017), "Moderators of intergroup discrimination in the minimal group paradigm: A meta-analysis," Duke University, mimeo.

Regner, T. and N. S. Harth (2014), "Testing belief-dependent models," Working Paper, Max Planck Institute of Economics, Jena.

Reuben, E., P. Sapienza and L. Zingales (2009), "Is mistrust self-fulfilling?" *Economic Letters*, 104: 89–91.

Ross, L., D. Greene and P. House (1977), "The "false consensus effect": An egocentric bias in social perception and attribution processes," *Journal of Experimental Social Psychology*, 13: 279–301.

Tajfel, H. (1970), "Experiments in intergroup discrimination," *Scientific American*, 223: 96–102.

Tajfel, H., M. G. Billig, R. P. Bundy and C. Flament (1971), "Social categorization and intergroup behavior," *European Journal of Social Psychology*, 1: 149–178.

Tajfel, H. (2010), *Social identity and intergroup relations*, New York, USA, Cambridge University Press.

Tajfel, H. and J. C. Turner, (1979), "An integrative theory of intergroup conflict," in W. G. Austin and S. Worchel (eds.), *The Social Psychology of Intergroup Relations*. Brooks/Cole, Monterey: 33–47.

Tajfel, H., and J. C. Turner (1986), "The social identity theory of intergroup behavior," in: Worchel, S., Austin, W. (Eds.), *Psychology of Intergroup Relations*, Nelson Hall, Chicago, pp. 7–24.

Vanberg, C. (2008), "Why do people keep their promises? An experimental test of two explanations," *Econometrica*, 76: 1467–1480.

Whitt, S. and R. K. Wilson (2007), "The dictator game, fairness and ethnicity in postwar Bosnia," *American Journal of Political Science*, 51: 655–668.

## Appendix A – Social identity and beliefs

In September 2017, we obtained in a five-session pilot experiment a significant correlation between behavior and beliefs conditional to the match. As shown in Table A, the second-order beliefs of the dictators who are matched with an out-group member are lower than those of the dictators who are matched with an in-group member (38% vs. 51%: $Z$=2.02, $p$=0.043) and the roll rates of the dictators who are matched with an out-group member are lower than those of the dictators who are matched with an in-group member (30% vs. 51%: $Z$=2.02, $p$=0.043). As mentioned in the main text, these correlations do not of course imply causation.

**Table A – Matching, second-order beliefs, and roll rates**

| | DICTATORS | |
|---|---|---|
| | SECOND-ORDER BELIEFS | AVERAGE ROLL RATES |
| MATCHING | (1) | (2) |
| (A) OUT GROUP | 0.38<br>(s.d. 0.32, obs. 356) | 0.30<br>(s.d. 0.46, obs. 356) |
| (B) IN GROUP | 0.51<br>(s.d. 0.33, obs. 412) | 0.51<br>(s.d. 0.51, obs. 412) |

## Appendix B – Elicitation of beliefs

The beliefs elicitation strategy is based on Vanberg (2008) and Di Bartolomeo *et al*. (2019a).

**Elicitation of first-order beliefs:** After matching groups, recipients were informed of the value of the switching probability in the round (they were aware that their paired subject could be switched according to that probability) and were asked to guess what their payoffs would be, i.e., what their (unknown) dictators would choose to do at the end of the round. Recipients could make their guess by ticking one of the five-point scale described in Table B. Beliefs were then re-scaled to 1, 0.75, 0.5, 0.25, and 0. The figures reported in the main text hence represent the averages of the recipients' re-scaled responses. The payoffs correspond to a quadratic scoring rule for probability values 85%, 68%, 50%, 32%, and 15% (under the assumption of risk neutrality, quadratic scoring yields flat payoffs as probabilities approach one).

**Table B – Incentives for first-order belief elicitation**

| The dictator will | choose *Roll* | | | choose *Don't Roll* | |
|---|---|---|---|---|---|
| | Certainly | Probably | Unsure | Probably | Certainly |
| Please tick your guess | ◯ | ◯ | ◯ | ◯ | ◯ |
| Your earnings if the dictator | | | | | |
| chooses *Roll* | 0.65 tokens | 0.60 tokens | 0.50 tokens | 0.35 tokens | 0.15 tokens |
| chooses *Don't Roll* | 0.15 tokens | 0.35 tokens | 0.50 tokens | 0.60 tokens | 0.65 tokens |

**Elicitation of second-order beliefs:** Before dictators were told whether their paired subject had been switched or not, they were asked to guess his guess. Specifically, they had to guess which of the five points of Table A had been ticked by their counterpart. Correct guesses were paid 0.50 tokens.

## Appendix C – First Round Results

Table C below contains the averages of dictators' second-order beliefs and *Roll* rates, the standard deviations and the number of observations reported in Table 3 of the main text, but considering only the first rounds. The rows of the table also distinguish: (i) dictators who were matched with an out-group member (row A) from (ii) dictators who were matched with an in-group member (row B).

**Table C. Matching, beliefs and *Roll* rates - first round only (192 obs.)**

| | DICTATORS | |
|---|---|---|
| | SECOND-ORDER BELIEFS | AVERAGE ROLL RATES |
| MATCHING GROUPS | (1) | (2) |
| (A) OUT GROUP | 0.43<br>(s.d. 0.29, obs. 96) | 0.32<br>(s.d. 0.47, obs. 96) |
| (B) IN GROUP | 0.50<br>(s.d. 0.33, obs. 96) | 0.54<br>(s.d. 0.50, obs. 96) |

Using the rank sum test, the *Roll* rate of in-group members (54%) is higher than that of out-group members (32%) [i.e., $z$=3.05, $p$=0.002]. The same result holds by using a one-tail difference mean test based on a $t$-test assuming dependent samples [i.e., $t$=3.12, $p$=0.001]. The average second-order beliefs of in-group members (50%) turn out to be higher than those of out-group members (43%) when a one-tail difference mean test is employed [i.e., $t$=1.69, $p$=0.047], but not when the rank sum test is used [i.e., $z$=1.60, $p$=0.111].[34]

[34] It is worth noting that the available observations are insufficient to reproduce the results summarized in tables 1 and 2 of the main text – and hence in tables 4 and 5 – by considering the first rounds only.