



ELSEVIER

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

Games and Economic Behavior

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

A stranger in a strange land: Promises and identity

 Gary Charness^{a,*}, Giovanni Di Bartolomeo^{b,c}, Stefano Papa^d
^a *University of California, Santa Barbara, United States*^b *Sapienza University of Rome, Italy*^c *University of Antwerp, Belgium*^d *University of Rome Tor Vergata, Italy*

ARTICLE INFO

JEL Classification:

A13

C91

D03

D64

D90

Keywords:

Social identity

In-group bias

Communication

Exogenous variation

ABSTRACT

One's social identity tends to favor those belonging to one's group. At the same time, communication has had beneficial social consequences in controlled laboratory experiments. Can communication improve trust and outcomes between out-group members by making them more familiar? We construct a simple weak mechanism of group favoritism (different assigned colors) that induces in-group favoritism. Communication increases cooperation but does not affect *per se* the social identity gap, i.e., people make more favorable choices for in-group members. However, mutual promises are more effective for out-groups than in-groups, reducing or eliminating the social-identity gap. So perhaps, in some sense, out-group members are seen as more like in-group members when they have experienced mutual promises. We also find evidence that the two groups' decisions are driven by different motivations, one based on expectations in-groups and one based on moral commitment in the case of out-groups.

1. Introduction

Experiments have shown that individuals tend to favor those who belong to their group (in-group bias). Often even artificial constructions of affiliation to a group are enough to produce in-group favoritism effects.¹ Tajfel and co-authors' groundbreaking studies were the first to demonstrate that individuals prefer rewarding in-group members over out-group members (Tajfel, 1970; Tajfel et al., 1971).² This tendency to maximize the outcomes of the in-group or to increase the differences between the in-group and out-group outcomes is a form of fundamental intergroup discrimination that can occur even without any meaningful group distinctions.

We investigate how the intersection of communication (mainly promises) and social identity affects cooperative behavior.

We thank the Journal Editor, Vincent P. Crawford, one Associate Editor, two anonymous reviewers, Giuseppe Attanasi, Martin Dufwenberg, and Kiryl Khalmetski, for their valuable comments. The hypotheses of the experiment were not pre-registered. However, our hypotheses, numbered H1–H5, were articulated before the experiment. Additional hypotheses were added during the review following some valuable comments from the referees. These should thus be considered exploratory.

* Corresponding author.

E-mail address: charness@econ.ucsb.edu (G. Charness).

¹ Even minimal-group assignments—or less, see Charness and Holder (2018)—can affect behavior (Chen and Li, 2009).

² These findings by Tajfel and his colleagues sparked significant interest in the study of social identity and paved the way for further research on intergroup bias and discrimination. Among those, Tajfel et al. (1979), Tajfel and Turner (1979, 1986). See also Tajfel (2010). Further references are provided below.

<https://doi.org/10.1016/j.geb.2023.12.005>

Received 19 July 2022;

Available online 23 December 2023

0899-8256/© 2023 The Author(s).

Published by Elsevier Inc.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Although there is evidence suggesting that the prominence of specific facets of identity can be shaped by environmental cues, offering a chance to promote a more favorable identity,³ an aspect that has received limited attention is the interplay between communication and social identity. Might communication also improve outcomes when people in relative out-groups interact? We designed and conducted a laboratory experiment to shed light on this question.⁴

We use a random-dictator game, augmented by pre-play communication and group membership, to fill this gap. Membership is induced by randomly assigning each participant to one of two groups (minimal-group paradigm), while communication consists of multiple bilateral messages. Our design adapts an idea from [Vanberg \(2008\)](#), providing a specific method to test the effects of communication using an exogenous variation. We study how a dictator who directly communicated with a recipient before their roles were revealed would have behaved, all else equal, in a context in which a dictator did not have communication with such a recipient. We do so by building a counterfactual based on a player-switching mechanism and asymmetric information, as we shall later explain.

Controlling beliefs, we test if (and when) communication and promises reinforce or weaken in-group favoritism. If there is an effect, it could go in either direction. On the one hand, communication could increase the salience of group distinctions (for example, imagine there are groups formed of partisans for one of two sports teams and the discussion centers on this sport). On the other hand, communication could reduce the social distance between people belonging to different groups or shift the focus to another facet of identity, thereby reducing any in-group favoritism.⁵ We also extend our investigation to the communication contents. By considering the messages, we can compare the relative effects of different kinds of communication (again using our exogenous variation). We focus on the effects of mutual promises that, in our bilateral-communication context, are mutual non-binding agreements to cooperate.⁶

We find that in-group favoritism is present. Subjects in-group matches sacrifice payoffs more frequently than dictators in out-group matches. One tends to choose cooperative actions more frequently without communication when paired with an in-group member. Direct communication between individuals increases cooperative behavior in all cases but does not *per se* reduce in-group favoritism. However, communication in the form of mutual promises does. Subjects are more likely to choose the favorable action if they have exchanged mutual promises with their matched partner. Within out-group pairs, the effect of mutual promises is strong enough to outweigh in-group favoritism, such that no favoritism is observed in the presence of promises. Our results suggest that in- and out-group promisors have different motivations for honoring mutual promises. Experimental outcomes support a guilt aversion (moral commitment) motivation for keeping a promise in the in-group (out-group) case. The difference drives an observed reduction in-group favoritism.

Our results also suggest that people care more about following the social norm when paired with an in-group member. People care about both their own monetary payoff and the appropriateness of norms in making their decisions. The impact of perceived social norms interacts with identity. The findings indicate that individuals respond to their perception of the social acceptability of their decisions and that this perception varies depending on whether the decision pertains to the in-group or the out-group. This was evidenced by the interplay between the data gathered from observed choices and survey responses. Experimental subjects care about the social appropriateness of their actions, and they do so differently when in-group or out-group. Agents are willing to pay 0.58 tokens to comply with social norms, with this willingness-to-pay increasing to 2.10 tokens when the subject interacts with an in-group member.

Our research is connected to two primary lines of literature. These encompass studies investigating social identity, notably the literature surrounding the minimal-group paradigm hypothesis, as well as research examining the impact of communication, particularly regarding promises. The most relevant papers are briefly discussed in the following section.

The remainder of this article is organized as follows. We present related literature in [Section 2](#). We discuss the experimental design, a sketch of the theory, and procedures in [Section 3](#). We discuss the messages sent and received in [Section 4](#). Our experimental results are shown in [Section 5](#). Finally, we provide a discussion and conclude in [Section 6](#).

2. Selected literature

[Tajfel and Turner \(1979\)](#) develop a theory of social identity, suggesting that people derive utility from group membership by behaving in accordance with group norms and stereotypes. However, the first consideration of identity in the economics literature is [Akerlof and Kranton \(2000\)](#), which proposes “a neoclassical utility function, where identity is associated with different social categories, each with a prescription or norm for behavior” ([Charness and Chen, 2020](#), p. 694). Deviations from these norms induce disutility. The framework has been applied in many economic areas, such as gender discrimination and the household division of labor.

[Kranton and Sanders \(2017\)](#) and [Kranton et al. \(2018\)](#) found that in-group bias exists in minimal and political group settings, as well as significant heterogeneity in behavior among individuals in group contexts. In dictator and trust games, various studies provide evidence for in-group favoritism. These include [Fershtman and Gneezy \(2001\)](#), [Ben-Ner et al. \(2009\)](#), [Whitt and Wilson \(2007\)](#), [Hargreaves Heap and Zizo \(2009\)](#), [Liebe and Tutic \(2010\)](#), [Ioannou et al. \(2015\)](#), and [Pan and Houser \(2013, 2019\)](#). Along the lines traced by [Chen and Li \(2009\)](#) conducted controlled experiments and found that in-group members exhibit more gift-giving, reciprocity, forgiveness, and overall efficiency than out-group members.

³ See [Shih et al. \(1999\)](#), [Chen et al. \(2014\)](#), [Charness et al. \(2014\)](#), and [Adnan et al. \(2022\)](#).

⁴ A series of laboratory experiments has shown that costless and non-binding communication (cheap talk) can effectively remove barriers to social efficiency. For example, [Charness and Dufwenberg \(2006\)](#) find that promises were especially effective in generating social efficiency.

⁵ [Ciccarone et al. \(2020\)](#) found in-group favoritism in the same context without communication.

⁶ [Feri et al. \(2010\)](#) and [Kocher et al. \(2020\)](#) discuss consensual institutions in team decision-making. See also [Di Bartolomeo et al. \(2023a,b\)](#).

We build on the papers focusing on the minimal-group paradigm. In a typical minimal-group experiment, subjects are randomly assigned to groups that are as meaningless as possible. The original minimal groups consisted of people who preferred paintings by either Klee or Kandinsky. Many experiments confirm and extend the [Tajfel and Billig \(1973\)](#) finding that group membership creates in-group enhancement in ways that favor the in-group at the expense of the out-group. Subjects in these experiments show that group membership affects individual behavior in treatments where groups are salient. Even with weak identity conditions, people tend to make choices that favor others identified as in-group members at the expense of out-group members.⁷

Several studies have also investigated the role of beliefs in-group favoritism. [Güth et al. \(2009\)](#) suggest that favoritism occurs because people dislike letting others' expectations down. [Daskalova \(2018\)](#) finds that expectations of favoritism lead to discrimination towards one's in-group in joint decisions, even if subjects do not discriminate individually. [Ockenfels and Werner \(2014\)](#) demonstrate that in-group favoritism may depend on beliefs since less favoritism emerged when dictators were told that the recipient was unaware of their shared group membership in a dictator game. [Ciccarone et al. \(2020\)](#) show a limited interaction between beliefs and social identity in a context without communication. Current evidence is thus mixed and probably depends on the context. Nevertheless, a key takeaway from the research is that we always need to check for possible effects through beliefs, which, inspired by these works, we do here.

The most closely related literature is the experimental research on pre-play communication in social dilemmas and similar settings. [Loomis \(1957,1959\)](#) and [Deutsch \(1958\)](#) study the impact of free-form written communication in simultaneous-move social dilemmas, finding that cooperation rates sometimes improve substantially when communication is allowed.⁸ [Ellingsen and Johannesson \(2004\)](#) and [Charness and Dufwenberg \(2006\)](#) study communication by the second player in settings with sequential moves. They find that written messages that take the form of promises are particularly helpful in generating trust and trustworthiness. For some recent contributions to this literature, see, e.g., [Vanberg \(2008\)](#), [Krupka et al. \(2017\)](#) and [Di Bartolomeo et al. \(2019b\)](#).

It turns out that the decision environment and the available communication technology are instrumental in determining the effect of this communication on choices and behavior. In some cases (e.g., [Cooper et al. 1989](#); [Charness 2000](#)), sending a simple message stating "I intend to play [A or B]" leads to a significant increase in social efficiency in coordination games (multiple equilibria), but did little where there is only one, socially inefficient equilibrium ([Charness, 2000](#)).⁹ Our communication technology is something of a hybrid, so it is not obvious *ex ante* what one might expect in our setting.

3. Experimental design, theory, hypotheses, and procedures

3.1. Experimental design

We are interested in social identity and communication. So, two crucial features of our design are group membership and pre-play communication (cheap talk). We build group membership by using a minimal-group paradigm. Each subject is randomly assigned to one of two groups: Red or Blue. By construction, our groups have low salience. We nevertheless expect to observe some in-group favoritism.¹⁰ A bilateral chat allows pre-play communication before a dictator game is played, as described below. As shall be explained, communication occurs before subjects know their role (dictator or recipient). For comparison, we investigate social identity in a dictator game as others have done (see, for example, [Güth et al. 2009](#); [Ockenfels and Werner 2014](#); [Ciccarone et al. 2020](#)).

Our experimental design is built around simple binary-choice random-dictator games, as seen in [Fig. 1](#). [Vanberg \(2008\)](#) introduced this idea, which has been used in many studies. Each dictator chooses between *Roll* and *Don't Roll*. Choosing *Don't Roll* leads to the dictator receiving 14 tokens and the recipient receiving nothing; choosing *Roll* leads to the dictator receiving 10 tokens, whereas the recipient receives 12 tokens with a probability of 5/6 and nothing with a probability of 1/6. We use payoffs from [Charness and Dufwenberg \(2006\)](#) to maintain some parallelism. These reflect monetary payments rather than utility since one's choice may be affected by social preferences (e.g., inequity aversion, reciprocity, guilt pangs). Who will become the dictator is initially unknown, but everyone has the same probability. We also assume that each recipient observes their own payoff but not the paired dictator's action.¹¹ The dictator and recipient may or may not belong to the same group.

The core of our design is a manipulation of the game described. After subjects are told their roles, half of the recipients are randomly re-matched with a new dictator (**partner switch**). The experiment thus involves switched agents matched with someone with whom they did not communicate and non-switched agents matched with the person with whom they communicated. The non-switched agents are thus "more familiar with" their partners than the switched agents are. Additionally, dictators know whether a switch

⁷ See, among others, See [Tajfel et al. \(1971\)](#) [Brewer \(1979\)](#), [Bourhis and Gagnon \(2001\)](#), [Charness et al. \(2007\)](#), [Charness and Villeval \(2009\)](#), [Chen and Li \(2009\)](#), [Hargreaves Heap and Zizzo \(2009\)](#), [Chen and Li \(2009\)](#), [Bénabou and Tirole \(2011\)](#), [Kranton and Sanders \(2017\)](#), [Kranton et al. \(2018\)](#), [Cason et al. \(2019\)](#), and [Ciccarone et al. \(2020\)](#).

⁸ Notably, the pioneering study of [Deutsch \(1958\)](#) considers the interaction between identity and communication, which is the main point of our paper. Thus, to our knowledge, Deutsch's work constitutes the birth of the investigation on how communication and identity have separate effects and how they potentially interact.

⁹ [Brandts et al. \(2016\)](#) and [Charness et al. \(2021\)](#) find considerable beneficial effects from chat in games but no effect in these same games from simple, check-a-box messages. See also [Charness and Dufwenberg \(2010\)](#) and [Di Bartolomeo et al. \(2019a\)](#).

¹⁰ We note that this format without communication was also successfully used by [Ciccarone et al. \(2020\)](#) to induce social identity. Other studies have also documented a favoritism bias induced by a minimal-group paradigm.

¹¹ In this paper, we adopt the recent gender convention of calling each individual "they."

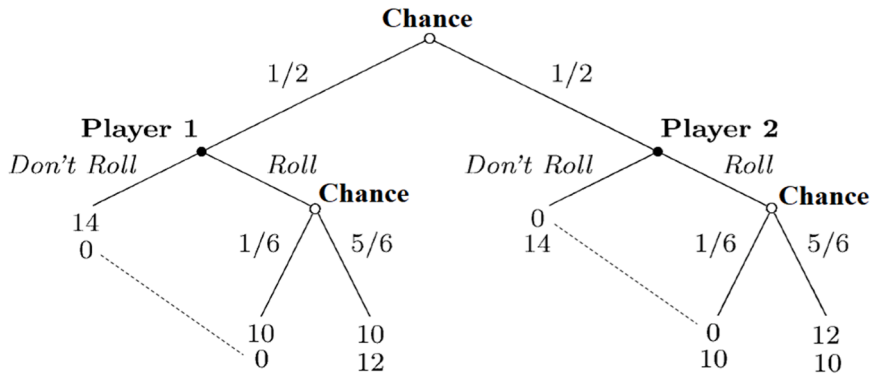


Fig. 1. A binary-choice random-dictator game.

has occurred while recipients do not, resulting in asymmetric information. If a switch occurs, the dictator a) observes the color of the new recipient and the color of the new recipient’s former match and b) reads the communication between their new recipient and the person with whom they were initially matched.

The experimental design is a special 2×2 factorial design (Blue/Red x Non-switched/Switched) of the simple random dictator game described in Fig. 1. To grasp the intuition, the switching mechanism at work is stylized in Fig. 2; this figure focuses, without loss of generality, on the case of Player 2 being selected as a Red dictator.

Before knowing their role in the game, Player 2 was initially involved in a chat with Player 1, and she could be either in an in-group match (left panel) or in an out-group match (right panel). After knowing their role, Player 2 is switched with a 50 % probability. If there is no switch, Player 2 is still matched with Player 1 (their initial partner) in an in-group (left panel) or an out-group match (right panel). Conversely, if switched, Player 2 can end up in an in-group or an out-group with the same probability, depending on whether they are re-matched with Player 3 or Player 4. In both cases, she is re-matched with someone who did not chat with them (i.e., a new partner).¹²

The partner-switching mechanism with asymmetric information implies that dictators know if they have been switched, but recipients do not. It means that the dictator’s beliefs about the beliefs of the recipient (second-order beliefs) are independent of the switching; this is a critical aspect of our design. Since each dictator is also informed about past communication and the color of the new partner, our design provides exogenous variations in both membership (good and bad matches) and communication (switched and non-switched). All else equal, according to the former, we can compare the behaviors in-group and out-group matches, assuming equal second-order beliefs. At the same time, the latter means that we can compare the behavior of dictators interacting with a new partner (switched) or the same partner (non-switched) given the same kind of content of communication (promises or not) with the same second-order beliefs.

Our interest is twofold. First, we are interested in studying direct communication’s effect on favoritism. So, we compare the behavior of non-switched subjects to comparable counterfactuals built on the switched ones. We also use the difference-in-difference between the in-group bias when a switch has occurred (i.e., no direct communication) and the in-group bias when no switch has occurred (i.e., direct communication.) Second, we investigate the effects of communication content on this in-group bias. We consider the relative effects of promises made to people who belong or do not belong to one’s own group. The first issue does not require a communication classification, while the second one does. The classification reflects mutual promises to act favorably. Recall that people do not know their roles when chatting.

3.2. Theory and testable outcomes

Akerlof and Kranton (2000) developed the first formal social identity model in economics.¹³ They propose a neoclassical utility function with identity associated with different social categories and a prescription for behavior. Prescriptions indicate the behavior appropriate for people in different social categories and situations, with deviations from the prescription leading to disutility.¹⁴ Shayo (2009) incorporates the choice of group identities and norms within a formal framework. He analyzes individuals’ decisions to identify with social groups, focusing on the effects of social status and social distance.¹⁵

Our paper hints at some theoretical ideas related to social identity and communication, and we build a simple pedagogical utility-based model to emphasize the mechanisms we have in mind and discuss possible complications and confounds. We also wish to

¹² Recall that, in both cases, Player 2 is informed of whether the new partner was initially in a good or bad match. This information is not described in the figure.

¹³ However, as stated, the idea is not new and has already been introduced by Tajfel (1970) and Tajfel et al. (1971) in the social psychology context.

¹⁴ Their versatile framework has been applied to gender discrimination, poverty and social exclusion, the household division of labor (Akerlof and Kranton, 2000), education (Akerlof and Kranton, 2002), and organization (Akerlof and Kranton, 2005).

¹⁵ See also Klor and Shayo (2010).

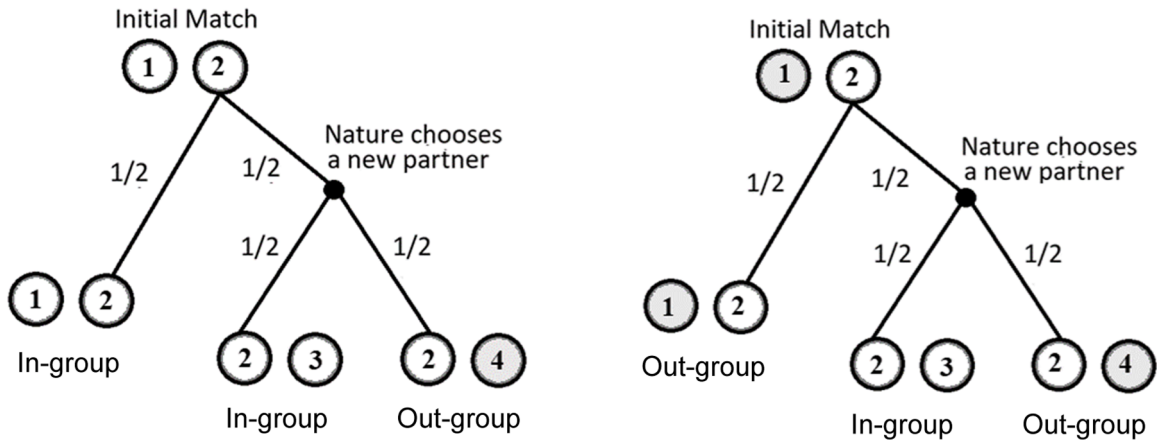


Fig. 2. The switching mechanism.

Note: The figure is adapted from [Ciccarone et al. \(2020\)](#). Numbers indicate the identity of hypothetical players; they are only shown to help the reader follow the sequence of matchings. The focus is on Player 2. The initial match refers to the original Player match 2 before the potential switch occurs. The switch is 50 % likely. If the switch occurs, independently of the kind of initial match, Player 2 can end up with an in-group or out-group match with a new partner.

highlight testable theoretical implications in our design. There are two dimensions whereby [Shayo \(2009\)](#) applies in our setting. First, the perceived social distance may be affected by the experience of communication. Second, a social group in which promises are featured may have higher social status in an individual’s eyes, thereby increasing the utility of belonging to the group or to the promisor partner.

3.2.1. A utility-based model

In the game described in [Fig. 1](#), if only monetary payoffs matter, no dictator will choose to roll since the dictator’s monetary payoff from choosing to roll is smaller. However, it is well-known (and observed empirically) that several factors imply that the dictator *i* could choose *Roll*. If we summarize these general motivations by α_i , the dictator *i* will choose *Roll* when $10 + \alpha_i > 14$, i.e., $\alpha_i > 4$.

In the following, we try to disentangle those motivations related to social identity and communication. To illustrate, we introduce step-by-step the relevant issues regarding our experiment. Henceforth, we introduce the following notation/indices $j \in \{I, O\}$, $h \in \{N, S\}$, $z \in \{M, A\}$ to identify dictators: *I* and *O* stand for In-group or Out-group,¹⁶ *N* and *S* stand for Non-switched or Switched, *M* and *A* stand for a dictator playing with a recipient who experienced a Mutual agreement or an Absence of a mutual agreement during the pre-play communication phase.

We will use the letters *j*, *h*, and *z* to denote the indices above described throughout the rest of this discussion. In addition, we will use $R()$ to indicate the average roll rates.

3.2.2. In-group favoritism

As discussed, extensive literature has shown that people favor their group members. [Akerlof and Kranton \(2000\)](#) use m_i to measure the impact of social identity on dictator *i*; in this vein, we can determine whether social-identity favoritism is activated by low-salience group membership, such as in our design. Specifically, if low-salience group membership results in social-identity favoritism, then $m_i(I)$ should be greater than $m_i(O)$.

Formally, augmenting our model with social identity entails that dictator *i* will roll if:

$$10 + m_i(j) + \alpha_i > 14 \quad j \in \{I, O\}$$

According to the above expression, when social identity is triggered, the likelihood of dictators rolling is higher if they belong to the in-group.

To examine the possibility of in-group favoritism, we analyze the roll rates of switched dictators as those are unaffected by direct communication by construction. We compare the average roll rate of switched dictators paired with in-group recipients to that of switched dictators paired with out-group recipients. Formally,

H1. In-group bias. Shared identity leads to higher roll rates: $R(I) > R(O)$ for switched dictators.

Drawing on the findings of [Ciccarone et al. \(2020\)](#) that revealed in-group favoritism in an environment bereft of communication, we expect H1 to hold. This can be explained by the fact that in-group members share a common identity, which leads to less social

¹⁶ By in-group and out-group, we refer to the match between the dictator and the recipient in the dictator game, who can be different from the match during the communication because of the switch.

distance between them.

3.2.3. Communication and social distance

It is widely believed that communication can decrease social distance. Therefore, the behavior of non-switched pairs may be influenced by their direct communication experience. Letting s_i represents the reduction of social distance from communicating together, we assume $s_i(N) > s_i(S) = 0$. In this case, the dictator i will roll if:

$$10 + m_i(j) + s_i(h) + \alpha_i > 14$$

The expression implies that the dictator i is more likely to roll when he directly communicates with the recipient—all else equal.

Although the hypothesis is slightly conjectural, there is considerable evidence from previous studies. H2 tests a straight communication effect, separate from identity considerations, i.e., $s_i(N) > s_i(S)$. In our design, since communication can have different effects for out-groups and in-groups, we can test two cases separately (in-group and out-group.)

H2. Straight effect of communication. Direct communication increases the roll rate: $R(N) > R(S)$ for both the in- and out-group subsamples.

H2 provides us with two tests: one for in-group dictators and another for out-group dictators.

We also assume that communication can interact with social identity. Our idea is that communication is a treatment that can “transform” out-group into in-group members; we assume that communication reduces the perceived distance between in-group and out-group members to something closer to that between two in-group members. So, $s_i(I, N) = s_i(N)$ and $s_i(O, N) = s_i(N) + \beta_i m_i(I)$, where β_i is the probability that communication transforms an out-group pair into an in-group pair.

Now, the dictator i will roll if:

$$10 + m_i(j) + s_i(j, h) + \alpha_i > 14$$

In our design, a testable implication for the above expression is that in-group bias within the non-switched dictators will be less than the bias within the switched ones. All else equal, in fact, $\beta_i > 0$ implies $s_i(I, N) - s_i(O, N) < s_i(I, S) - s_i(O, S)$. H3 follows:

H3. Effect of communication on in-group favoritism. Direct communication reduces the in-group bias. Formally, $R(I, N) - R(O, N) < R(I, S) - R(O, S)$.

The above difference-in-difference comparison captures the effect of the treatment on in-group bias. As formalized, we presume that communication will “transform out-group into in-group members,” thereby reducing any difference found without it. Again, in principle, this could go the other way since people might comment on group differences, reinforcing these.

3.2.4. The interaction between communication contents and social identity

We investigate the mere effects of communication by comparing non-switched and switched dictators. Nevertheless, communication can be further qualified by inspecting its contents. In this respect, we expect that promises could have specific effects on in-group bias. Classifying communication content involves some subjectivity, making the investigation more challenging. Section 4 presents our classification strategy. Here we describe the hypotheses that we consider.

We focus on mutual promises (agreements), meaning that each party assures the other party of their intention to roll as a would-be dictator.¹⁷ Breaking one’s promise has a cost. Hence, whoever makes a promise to the recipient is more likely to choose to roll. Defining the cost of breaking a promise by $c_i > 0$. Now, the dictator will roll if:

$$10 + m_i(j) + s_i(j, h) + \alpha_i > 14 - c_i(j, z)$$

The above expression implies that we expect that promisors are more likely to roll, as breaking a promise has a cost $c_i(j, M)$, while there is no cost when no promise has been made, $c_i(j, A) = 0$.

Do promises interact with social identity? We might expect in-group bias to decrease with promises to the extent that they can substitute for in-group membership. The perceived social distance between individuals was small for in-group pairs but more substantial for out-group pairs. Mutually favorable assurances could close this gap.

H4. Effect of promises on in-group bias. Agreements are more effective in the out-group matches than in the in-group matches: $R(O, N, M) - R(O, N, A) > R(I, N, M) - R(I, N, A)$.

3.2.5. Promise-keeping motivations and social distance

As our discussion revolves around fulfilling promises, we also explore the reasons behind this behavior. We highlight two possible justifications: There might be a moral obligation where individuals honor their promises regardless of what others may expect (Vanberg, 2008). Alternatively, a second explanation is based on expectations: promises can influence people’s beliefs, and individuals may be averse to guilt from letting down someone’s expectations.

¹⁷ We hypothesize that the chat’s impact will be more pronounced if we focus solely on promises. This prediction is based on the belief that direct promises inherently reduce the perceived social distance for out-group members (and potentially increase the in-group’s status) while having little or no impact on the perceptions of in-group members. A more detailed analysis of communication contents is provided in the on-line appendix.

Both explanations are consistent with an observed correlation between promise-keeping and second-order beliefs. Vanberg (2008) thus proposes a simple, innovative test for moral commitment based on the same information structure used in our paper. He compares the behavior of dictators who could keep their promises to that of those (switched) who could keep promises made by others. The asymmetric information design implies that switched and non-switched dictators should have the same second-order beliefs, all else equal. Therefore, evidence for moral commitment is provided when dictators only keep their own promises; he finds such evidence.

We use Vanberg's test to explore the motivation for promise-keeping within in-group (or out-group) pairs. In keeping with Vanberg (2008), although in a different context, we suspect that moral commitment affects dictator behavior. The social-identity models do not generate the following hypotheses, although they are consistent with them.

In our setup, second-order beliefs of switched and non-switched dictators are independent of their switching condition since recipients cannot observe switching. Hence, we can compare the behavior of dictators who made pledges to dictators asked to keep pledges made by others. We can separate the cost of not honoring a promise into two components: $c_i = c_i^m + c_i^g$. The former captures the cost of moral commitment, while the latter captures the cost from belief-dependent motivations. Our information structure implies that $c_i^g(N) = c_i^g(S)$ by construction, since beliefs are independent of the switching condition, so moral commitment then implies $c_i(N) > c_i(S)$. In this respect, a test for moral-commitment-driven motivation in promise-keeping follows.¹⁸

H5. Moral commitment (Vanberg's) test. People are less likely to keep mutual promises made by others than by themselves: $R(N, M) > R(S, M)$ for in- and out-group dictators.

H5 provides us with two tests: one for in-group and another for out-group dictators.

Of course, communication dynamics can be more complex than the simple dichotomy of promise and no-promise. So, we also considered the effects of communication from a broader perspective, i.e., how the roll rates differ across different sub-categories that do not involve initial promises. However, using finely grained categories for communication content reduces statistical power when comparing message sub-classifications. Thus, we relegate such additional results to the online appendix with a caveat: these results should be seen as suggestive rather than conclusive.

3.3. Procedures

The experiment was conducted at the Sapienza University of Rome (CIMEO Lab). We used a within-subject design that involved 384 students (12 sessions, 10 rounds each, 32 subjects each) recruited using an online system. Upon arrival, subjects were randomly assigned to 32 isolated computer terminals.¹⁹ Two assistants handed out instructions and checked that participants correctly followed the procedures. The experiment took place between April and May of 2021.

The experimental design consisted of three stages: 1) group assignment (pre-session), where pairs of subjects observed their colors; 2) experimental session; 3) final payment (post-session). One's color remains the same throughout the experiment. In each of the 10 rounds, subjects interact with a different participant chosen randomly by computer. In no case does any subject interact with the same participant twice (perfect stranger matching). After the group assignment, the session continued. Each round implemented the following sequence of five stages:

- (1) *Communication.* Subjects were randomly matched to form 16 pairs, with a random determination of who would start the chat. Each chat consisted of up to four one-way posts in sequence, each of which could have up to 90 characters.
- (2) *Role assignment.* At the beginning of the round, roles (dictator or recipient) were randomly assigned for each pair, and subjects were so informed.
- (3) *Switching.* Half of the pairs were switched. Only dictators were informed of whether a switch occurred. Each switched dictator with a new recipient learned: a) the color of the new recipient, b) the color of the previously-paired dictator of the new recipient, and c) the prior conversation that the new recipient had with the previously-paired dictator.
- (4) *Belief elicitation.* This stage has two parts: a) first-order beliefs: each recipient was asked to guess if the paired dictator (after any switch) would choose to *Roll*, and b) second-order beliefs: dictators, when there was a switch, were asked to guess the guess of the current recipient, for both those who were switched and those who were not.²⁰
- (5) *Dictator's action.* The dictator chose *Roll* or *Do not Roll*. All subjects were informed of their payoff for the round. When the received payoff was zero, a recipient could not be sure of the dictator's choice.²¹

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 Fig. 1 were computed in tokens (where 1 token = 0.5 euro). Subjects were told

¹⁸ Note also that a fair comparison requires considering only the subsample of switched dictators who made promises during the pre-play phase.

¹⁹ The experiment was designed using z-Tree (Fischbacher, 2007).

²⁰ Belief elicitation is described in detail in the next section. We stress that we elicit beliefs after dictators know whether they have been switched before dictators make their choice. Hence, we elicit the beliefs of switched and non-switched dictators' beliefs regarding their recipients' beliefs, as in Di Bartolomeo et al. (2019b, 2023b) and Vanberg (2008). Since switched dictators should infer the beliefs of the new recipients, elicitation is done after they read the chat of their partners. Again, dictators know that recipients do not know if they were switched.

²¹ Recipients could obtain a zero payoff in one of two ways: (i) their dictator had chosen *Don't Roll* or (ii) their dictator had chosen *Roll* and the outcome of the die-roll was "1."

truthfully that incentives for belief elicitation were provided for all rounds except the one chosen for payment; this removed any incentive to hedge by misreporting beliefs).²²

4. Messages

Communication between participants is at the heart of our experiment. Research assistants blind to our hypotheses cataloged the messages.²³ Recall that communication is bilateral and has back-and-forth messages. Moreover, participants chat before learning their role, knowing that one in each pair will be a dictator and the other will be a recipient.

We note that this protocol differs somewhat from those used previously. Charness and Dufwenberg (2006) used free-form chat with a one-page limit and only one message sent from the second to the first mover. While we might expect more effective communication with more communication rounds, the 90-character restriction on each message could limit this. In our communication protocol, participants may be inclined and able to strike a deal of conditional cooperation: “I’ll pledge to choose *Roll*, if you also pledge to choose *Roll*.” When both players do so, we find it natural to define a “promise” as a promise to keep this form of agreement.²⁴ Operationally, we asked the research assistants to classify each subject’s message according to whether it conveyed a pledge.²⁵ We observe a “mutual promise (agreement)” if and only if both sides state their intention to roll (*RR*), where *R* refers to the pledge. By contrast, we classify all other communication outcomes as non-promises (absence of a mutual agreement). Our assumption is substantial since the communication dynamics may also affect behaviors. A dictator can infer some information about the kind of recipient she faces from reading his communication. In principle, that information can be relevant. For instance, she could be more likely to support a recipient claiming to be ready to roll.²⁶

Some examples may help to clarify.²⁷ We present these in order of the intuitive chance the dictator will subsequently choose to roll. Either player could become the dictator at this point. Examples 1a and b involve Mutual agreement (*M*), while the others contain no agreement to roll, i.e., the absence of a mutual agreement (*A*).

Example 1a. Both people state that they would like to roll. (*RR*)

Player No 1 says “If I am Player A [dictator], I will roll the dice.”

Player No 2 then replies, “I would roll the dice, too.”

Player No 1 says “So, let’s roll, Player A [dictator] will roll!”

Player No 2 says “OK!”

Example 1b. (*RR*)

Player No 1 says “do you pledge to roll the die?”

Player No 2 then replies, “Yes, you too, though.”

Player No 1 says “I pledge!”

Player No 2 says “OK!”

Example 2a. The dictator pledges to roll while the recipient did not. (*RN*)

Player No 1 (future dictator) says “I will roll!”

Player No 2 replies “I am not sure.”

Example 2b. (*RN*)

Player No 2 says: “New washing machine launched on the market: 12 dead, 4 injured. Will you roll the die?”

Player No 1 (future dictator) replies “I always roll the die.”

Player No 2 says “You like to live on the edge, huh?”

Player No 1 replies “Maybe.”

Example 3a. The dictator made no cooperative statement, but the recipient did (*NR*)

Player No 2 says “I will roll!”

Player No 1 (future dictator) replies “I am not sure.”

²² A reader mentioned that we could have instead studied how concerns related to group identity affect the formation of trust and cooperation. Indeed, this is a key topic. However, our experimental design is not geared to this question, so this would be a useful and related additional paper.

²³ The research assistant was not involved in the design and execution of the experiment. Indeed, we asked three research assistants to classify messages, and we *ex-ante* randomly chose the classification of one of them for the experiment. The different classifications were strongly correlated (with a high Cronbach alpha value of 0.888.)

²⁴ This random-dictator design is, therefore, different from Charness and Dufwenberg (2006). On this point, see Di Bartolomeo et al. (2023b), where unilateral one-shot communication is instead considered.

²⁵ As in Vanberg’s protocol, each pair of chat messages sent by the same subject in a round was treated as a unit.

²⁶ We look at communication dynamics in an online appendix, where we consider a more detailed classification for other messages than agreement: “dictator’s unilateral promise,” “recipient’s unilateral promise,” and “no promise at all.” However, controlling for all taxonomies reduces the power of the tests since the number of observations in each resulting case drops correspondingly.

²⁷ All the examples are from the experiment chats (translated from Italian.)

Example 3b. (NR)

Player No 2 says “I roll the die.”
 Player No 1 (future dictator) replies “I will not.”
 Player No 2 says “alright!!!”
 Player No 1 replies “ok.”

Example 4a. (NN)

Player No 1 says “Life is an abyss.”
 Player No 2 replies “Yes but that’s how life goes, I’m a constant carpe diem; so, I don’t waste time and I live.”
 Player No 1 says “Very well”
 Player No 2 replies “While you observe the abyss, the abyss observes you; if you roll the dice, I’ll give you a rope, so you don’t fall.”

Example 4b. Neither person disclose a positive intent. (NN)

Player No 1 says “I will not roll the dice.”
 Player No 2 replies “Me neither.”

Our experiment involves 384 subjects playing 10 rounds. Thus, our sample consists of 3840 messages (1920 chats), shown in [Table 1](#). Among the chats, 1113 (58 %) were considered mutual agreements.

It is worth noting that messages considered here include both dictators and recipients since chats occurred before subjects knew their role. With in-group matches, 605 of 1008 chats (60.0 %) were mutual agreements; with out-group matches, 508 out of 912 (55.7 %) were mutual agreements. [Table 1](#) shows the details, i.e., mutual agreement rates made with out-groups and in-groups in each session. The rate of mutual agreements was slightly higher for in-group matches but not significantly so ($Z = 1.18, p = 0.239$.)

5. Results

We present our results in three steps. First, we focus on direct communication and social identity (H1-H3). Second, we examine the relationship between promises and social identity (H4). Third, we look at motivations (H5); here, we explore data by considering how promises may change the perception of group identity using survey-style measures of shared identity.²⁸ Our design involves several exogenous variations. Before discussing the actions chosen by the participants, we will always discuss and test these variations, which involves considering the elicited second-order beliefs. Our non-parametric analysis uses Wilcoxon signed-ranks tests on session-level data throughout the paper. We conservatively report two-tailed tests.²⁹ We present the session averages in the main text for a more straightforward exposition, while the detailed session rates are reported in the appendix.

5.1. Direct communication and social identity

[Table 2](#) displays the mean second-order beliefs and *Roll* rates by identity (in-group or out-group) and switching condition (switched or not). The average second-order beliefs are also presented to test exogenous variations.

Before examining the behavior of dictators, we begin by testing for exogenous variations. These imply the same average second-order beliefs of switched and non-switched dictators in any columns. One would hope this is true since dictators know that the recipients do not know if they have been switched. All else equal, the table shows no statistically-significant differences in the second-order beliefs of switched and non-switched dictators for both out-group cases (68 vs. 68 %) and in-group cases (67 vs. 70 %). We find $Z = 0.39, p = 0.695$ across out-groups and $Z = -1.69, p = 0.091$ across in-groups. The table also shows two other exogenous variations in group membership. There are no differences in the second-order beliefs of out-group and in-group dictators within either non-switched subjects (68 vs. 70 %) or within switched subjects (68 vs. 67 %).³⁰

Now let us look at the roll rates, i.e., the switched dictators’ behavior. The horizontal comparison (1) provides evidence in favor of in-group favoritism among dictators who did not directly communicate, as expected. The average *Roll* rate with in-group matches (50 %) is significantly higher than with out-group matches (40 %), with $Z = 2.94, p = 0.003$. Result 1 follows.

Result 1 (favoritism). Data support in-group favoritism (H1): In-group dictators who did not directly communicate with their in-group counterparts are more likely to roll than out-group dictators who did not directly communicate with their out-group counterparts.³¹

Does having had direct communication affect the behavior of dictators? To answer this question, we compare the behavior of dictators who communicated directly (non-switched) with those who did not (switched). So, we compare the roll rates across columns (b.1) and (b.2) in [Table 2](#) (H2). Comparing out-group switched and non-switched dictators (b.1), the average *Roll* rate increases from

²⁸ This analysis has been added ex-post (we thank a referee for suggesting it.)

²⁹ All tests are two-tailed Wilcoxon signed-ranks tests on session-level data unless otherwise specified. All p -values are rounded to the nearest three decimal places.

³⁰ We find $Z = -0.94, p = 0.347$ for non-switched and $Z = 0.98, p = 0.327$ for switched.

³¹ Similarly, the horizontal comparison (2) provides evidence in favor of in-group favoritism with non-switched dictators (those who did directly communicate). The 58% average *Roll* rate of in-group dictators is statistically higher than the 50% average *Roll* rate of out-group dictators ($Z = 2.04, p = 0.041$). Hence, favoritism is indeed observed among both switched and non-switched dictators.

Table 1
Matching and mutual agreements by session (1920 obs.).

Category	Session												All
	1	2	3	4	5	6	7	8	9	10	11	12	
Out-group	.62	.54	.54	.58	.78	.61	.49	.59	.21	.47	.67	.59	.56
obs.	76	76	76	76	76	76	76	76	76	76	76	76	912
In-group	.58	.60	.67	.67	.77	.73	.54	.71	.45	.38	.52	.58	.60
obs.	84	84	84	84	84	84	84	84	84	84	84	84	1008

Table 2
Matching, second order beliefs, and roll rates (1920 obs.).

	SOB		Roll	
	Out-group	In-group	Out-group	In-group
(1) Switched	(a.1) .68 (468)	(a.2) .67 (492)	(b.1) .40 (468)	(b.2) .50 (420)
(2) No switched	.68 (420)	.70 (540)	.50 (492)	.58 (540)

40 to 50 %, while, with in-group dictators (b.2), this rate increases from 50 to 58 %. In both cases, differences are statistically significant with conservative tests ($Z = 2.75, p = 0.006$ and $Z = 2.35, p = 0.019$, respectively).³² It follows that direct communication matters.

However, does it affect in-group bias? Does direct communication affect in-group dictators more than those in out-group matches? We use our difference-in-difference approach to answer this question (H3). The average Roll rate for out-group dictators increases from 0.40 to 0.50 with the treatment (communication among non-switched pairs), while it increases from 0.50 to 0.58 for in-group dictators. However, there is no significant difference between the differences (0.10 vs. 0.08; $Z = 0.94, p = 0.347$).

Outcomes from H2 and H3 testing can be summarized in the following result.

Result 2 (direct communication). Our findings indicate that dictators who stay within the same group and experience direct communication tend to exhibit higher average roll rates, with comparable effect sizes observed for both in-group and out-group dictators. Hence, direct communication matters but does not affect in-group bias, having the same impact on in-group and out-group dictators.

5.2. Promises and social identity

To expand upon the findings from the previous section, we examine the communication content. Table 3 displays the mean second-order beliefs and Roll rates of non-switched dictators who either made an initial mutual promise or did not and switched dictators who made an initial promise and were later paired with one who did not receive a promise. The table also distinguishes between in-group or out-group pairs. The average second-order beliefs (SOBs) are also presented to test exogenous variations.

Again, before examining the behavior of dictators, let us begin by testing for exogenous variations; we consider two additional variations related to group membership. In this respect, we need to focus on second-order beliefs. Columns (a.1) and (a.2) display second-order beliefs of non-switched dictators who did or did not make an agreement, respectively, while column (a.3) shows those of switched dictators who agreed to roll and were paired with someone who did not. The table also distinguishes between the out-group and in-group scenarios.

No significant differences were found between the second-order beliefs of in-group and out-group subjects in any of the three cases (again, note that recipients did not observe the switch).³³ The table also reveals no difference in the second-order beliefs of 1) switched dictators involved in a mutual promise who were paired with someone who did not receive a promise and 2) non-switched dictators who were not involved in a mutual promise.³⁴ Exogenous variations are thus successfully tested. We note for future reference that the table shows that individuals who experienced mutual promises have higher second-order beliefs than those who did not.³⁵ This finding aligns with the well-known correlation between promises and second-order beliefs (Charness and Dufwenberg, 2006).

³² The session-level data (see the appendix) show comparing across switched (non-switched) categories.

³³ Formally, for non-switched subjects not involved in agreements (a.1), the second-order beliefs were 51 vs. 55 %, with a Z score of -0.63 and a p-value of 0.530. Similarly, in scenario (a.2), the second-order beliefs were 81 vs. 80 %, with a Z score of 0.55 and a p-value of 0.583. Lastly, no significant difference was observed for switched dictators who made a mutual promise and were paired with someone who did not receive a promise (a.3), with second-order beliefs of 53 vs. 56 %, a Z score of -1.02 , and a p-value of 0.308.

³⁴ No difference is observed in either the in-group sample (0.56 vs. 0.55; $Z = 0.75, p = 0.456$) or the out-group sample (0.53 vs. 0.51; $Z = 1.02, p = 0.308$).

³⁵ In detail, 0.80 vs. 0.55 for the in-group sample and 0.81 vs. 0.51 for the out-group sample: $Z = 3.06, p = 0.002$.

Table 3
Matching, SOBs, and rolls for selected dictators' categories (1169 obs.).

Dictator → Recipient →	SOBs			Roll rates		
	not switched		switched	not switched		switched
	no promise no promise	promise promise	promise no promise	no promise no promise	promise promise	promise no promise
	(a.1)	(a.2)	(a.3)	(b.1)	(b.2)	(b.3)
(1) Out-group	.51 (187)	.81 (233)	.53 (96)	.28 (187)	.68 (233)	.30 (96)
(2) In-group	.55 (209)	.80 (331)	.56 (113)	.39 (209)	.69 (331)	.50 (113)

Note: Rows indicate the out-group (1) or in-group members (2). Column letters indicate average SOBs (a) and Roll rates (b). The column sub-numbers indicate the different selected cases: (0.1) stands for non-switched dictators not involved in mutual promises during pre-play communication; (0.2) stands for non-switched dictators involved; (0.3) stands for switched dictators who made a mutual promise and were paired with someone who did not make a promise.

After exogenous variations have been tested, we can focus on dictators' behavior, i.e., columns (b.1) - (b.3). Looking at the roll rates from a top-down perspective, we see that non-switched dictators who were not involved with mutual promises were more inclined to sacrifice payoffs when paired with in-group recipients as opposed to out-group recipients (0.39 vs. 0.28: $Z = 2.20$, $p = 0.028$). In contrast, there is no discernible difference when considering those (in-group vs. out-group) who have mutually agreed to roll (0.69 vs. 0.68, $Z = 0.75$, $p = 0.456$).

Do our data suggest that promises eliminate the in-group bias? People who made promises were much more likely to sacrifice than those who made no promises for in-groups (0.68 vs. 0.28: $Z = 3.06$, $p = 0.002$) and out-groups (0.69 vs. 0.39: $Z = 3.06$, $p = 0.002$). As seen above, there is no difference between in-group (0.68) and out-group promises (0.69), with $Z = 0.75$, $p = 0.456$). Summing up, we find an in-group bias among non-switched dictators without promises. However, there is no in-group bias amongst non-switched dictators involved in an initial mutual promise.

However, the above comparison was made between subjects of two sets: one group involved in an initial mutual promise and the other without such a promise. As these groups could consist of individuals with different characteristics, a fairer comparison would be to assess how the dictators in the mutual promise group would have acted had they not made such a promise. To achieve this, we can construct a counterfactual in our design by utilizing the data generated by our switching mechanism.³⁶

In this respect, while experimental outcomes show no group bias for non-switched dictators who made a mutual agreement, the comparison in column (b.3) indicates a substantial and significant in-group bias (0.50 vs. 0.30: $Z = 2.20$, $p = 0.028$) for the counterfactual of switched dictators involved in a mutual promise but later re-matched with a recipient who was not.³⁷

Result 3. Promises effectively eliminate the in-group bias otherwise observed.

This finding requires further examination since it suggests mutual agreements are more effective for individuals with out-groups. The difference-in-difference comparison in Fig. 3 shows in-group bias by calculating the difference between the average *Roll* rate of in-group and out-group dictators. Specifically, we compare the in-group bias for non-switched dictators who experienced a mutual promise with that of counterfactual switched dictators who made a mutual promise and were re-matched with recipients who did not experience a mutual promise. The difference between the dark and gray points on the promise line quantifies the former, while the latter is determined by the distance between the dark and gray points on the no-promise line.

The difference-in-difference comparison provides evidence that promises are more effective for the out-group people (i.e., $0.68 - 0.30 = 0.38$, while $0.69 - 0.50 = 0.19$, with $Z = 1.96$, $p = 0.049$), so not only do promises significantly increase *Roll* rates, but they also eliminate in-group bias. In the no-promise case, out-group dictators are less likely to roll than the in-group ones (despite having the same second-order beliefs). After being involved in a promise, all dictators are more likely to *Roll*, but a promise is more effective when the counterfactual situation is an out-group recipient (again, although second-order beliefs are the same.)

Similarly, we can test a promise's impact on in- and out-group pairs. Fig. 3(a) compares individuals who have made a promise to those who have made one but find themselves in a counterfactual situation where no promise was made (i.e., 0.20 vs. 0.01: $Z = 1.96$, $p = 0.049$).

Result 4. Overall, we support H4, i.e., mutual promises decrease favoritism. Promises to *Roll* do not affect behavior independently from the in-group or out-group status – they are more effective in the latter case and tend to reduce or even eliminate in-group bias.

5.3. Exploring motivations

We now explore the rationale for promise-keeping, distinguishing motivations of out-group dictators from those of in-group

³⁶ We also note that people who made a favorable declaration may have different attitudes than those who did not. In-group who experienced a promise and (once switched) read a non-promise message were more likely to roll than in-group people who did not experience a promise (0.50 vs. 0.39: $Z = 2.43$, $p = 0.015$). However, there is no statistically significant difference in the out-group subsample (0.30 vs. 0.28: $Z = 0.86$, $p = 0.388$).

³⁷ The session-level data (reported in the appendix) provide further information.

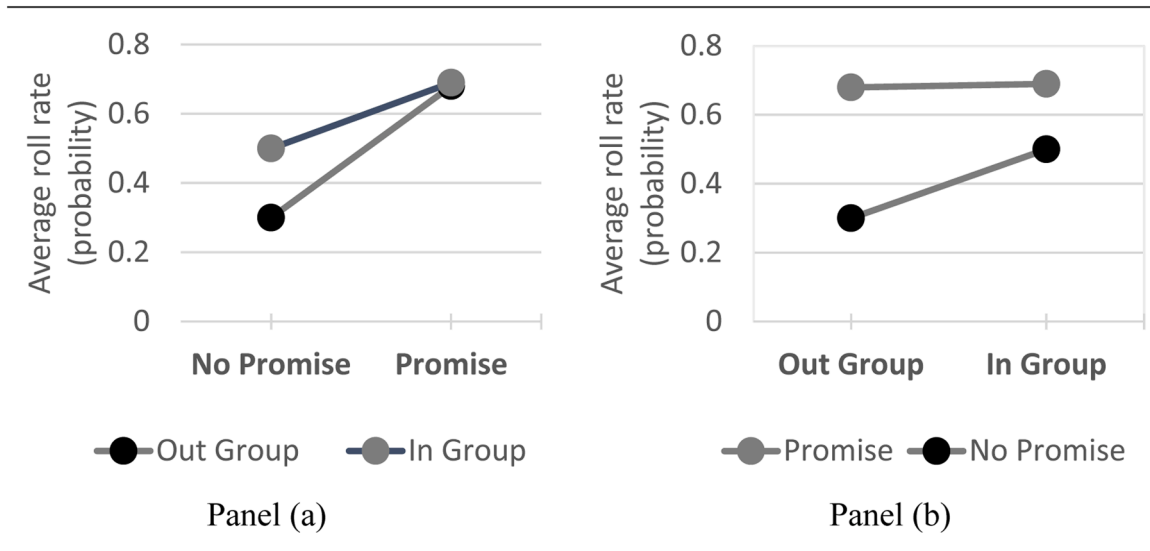


Fig. 3. The effects of making a promise on in-group favoritism.

dictators. In a nutshell, we apply Vanberg's test (H5) within the two sub-samples to separately test their motivations. We focus on dictators who made a promise, comparing the behavior of non-switched dictators to those dictators re-matched with a recipient who has received a promise from another. To compare second-order beliefs fairly, we account for the dictators re-matched with recipients previously out-group (in-group) when the out-group (in-group) comparison is considered.³⁸

The Roll rates are provided in Table 4. At the top of the table, we focus on out-group dictators, while at the bottom, we look at in-group dictators. The table reports the roll rates of the non-switched dictators and those of the switched ones for recipients who were initially out- or in-group. Again, a fair comparison requires that, e.g., non-switched out-group dictators who made a promise (row (1), column (a)) should be compared to the switched out-group dictators who made a promise and played with a recipient who initially was out-group and received a promise (row (1), column (b)). Similarly, we compare columns (a) vs. (c) of row (2) for in-group dictators.

We test moral commitment (H5) by comparing (a) with (b) in the first row and (a) with (c) in the second row since second-order beliefs are the same for switched and non-switched dictators within the out-group and in-group samples, as we have argued that they should be. Our results provide evidence that the behavior of out-group promisors is consistent with a moral commitment, i.e., they are significantly more likely to keep their own promises (68 vs. 53 %: $Z = 2.58$, $p = 0.009$). H5 receives support for out-group dictators. Conversely, the behavior of in-group dictators does not show evidence of moral commitment: all else equal, they also keep promises made by others (69 vs. 69 %: $Z = 1.09$, $p = 0.272$). H5 is thus not supported for in-group dictators. Their behavior is consistent with choices driven only by a sense of guilt.

Result 5. The motivation for keeping promises differs somewhat between in-group and out-group promisors. Out-group promisors appear to be more motivated by moral commitment, whereas in-group promisors are more motivated by a desire to avoid feelings of guilt.

Exploring the data further by considering comparisons that allow us to investigate the interaction between promise-keeping and social identity is also helpful. For example, we can test if a switched promisor dictator is willing to keep a promise made by a member of his group to another group member. Similarly, we can also look at the case where a switched promisor dictator is willing to keep a promise made by another group member to a member of his group. Specifically, the comparisons between (a)–(c) for the out-group members (i.e., row (1)) and (a) and (b) for the in-group members (i.e., row (2)) provide us with two outcomes.

Comparisons lead to the following outcomes.

- (1) Out-group promisors are more likely (68 vs. 52 %) to keep their own promises than dictators who play with an out-group member who previously received a promise from an in-group member, but this effect is not statistically significant ($Z = 1.33$, $p = 0.182$).
- (2) In-group promisors are more likely (69 vs. 59 %) to keep their own promises than dictators who play with an out-group member who previously received a promise from an in-group member, but this effect is not statistically significant ($Z = 1.41$, $p = 0.158$).

³⁸ We verify the exogenous variation by testing the equality in the second-order beliefs. In-group promisors who read a promise by a new ingroup recipient have the same second-order beliefs as the ingroup who made promises (0.80 vs. 0.80: $Z = 0.31$, $p = 0.754$). Outgroup promisors who read a promise by a new outgroup recipient have the same second-order beliefs as outgroup individuals who made promises (0.83 vs. 0.81: $Z = 0.18$, $p = 0.859$). See the online appendix.

Table 4
Average Roll rates of selected dictators involved in a mutual promise (718 obs.).

	Non-Switched	Switched and re-matched with a recipient also involved in a promise and was originally in an	
	(a)	out-group match (b)	in-group match (c)
(1) Out-group	.68 (233)	.53 (77)	.52 (94)
(2) In-group	.69 (331)	.59 (92)	.69 (77)

Note: Rows indicate final matches; out-group (1) or in-group members (2). Columns indicate the initial matches of recipients. Of course, in the case of non-switched (a), these are the same as the dictators, while in the case of switch, they might have been out- (b) or in-group (c). All dictators and recipients used to build the table were involved in a mutual promise during the pre-play communication.

Combining our results, we summarize the interaction between promise-keeping and social identity: A dictator is likely to keep their own promises to any co-player independently of the group to which they belong. A switched dictator is also likely to keep a promise made by or to someone from their group. By contrast, a switched dictator is less likely to keep promises made by someone who does not belong to their group.³⁹ These outcomes support the idea that the shared in-group identity with the other promise-maker also makes the dictator feel somewhat bound by that promise.⁴⁰ However, although promises reduce the effects of in-group favoritism, they do not fully eliminate them since dictators of a specific group less frequently keep promises made by dictators from a different group to recipients from a different group.

As shown in the previous sections, reducing in-group favoritism reflects mutual promises involving out-group subjects being relatively more effective than those of in-group subjects. So, our result on motivation implies that the reduction of in-group favoritism is thus mainly driven by the moral-commitment effect that characterizes the motivations of out-group dictators. A difference-in-difference approach offers much the same result. We compare the 15 percentage-point increase in own promise-keeping in the out-group sample to the 0 % difference in the in-group sample, finding a statistically significant difference-in-difference ($Z = 2.31$, $p = 0.021$). So, when asked to honor a pledge, the increase in promise-keeping is significantly more significant for out-group dictators than for in-group ones.

We further explore the argument about promises changing the perception of group identity by using survey-style measures of shared identity.⁴¹ We adopt Krupka-Weber's method and investigate the issue in terms of social norms. In a nutshell, we apply a social norm to a possible binary action ("roll after you made a promise" or "do not roll after you did not make a promise.") The two actions lead to the same outcomes, but they can vary in the degree to which they are perceived as socially appropriate. Hence, observed choices may be governed by different social norms deriving from different group identities.⁴²

Survey outcomes are reported in Table 5. Data show that keeping a promise is always perceived as socially appropriate for both in-groups and out-groups. Breaking a promise is considered most socially inappropriate (56.6 %) when an in-group member is involved, but many people (41.5 %) felt it was nevertheless somewhat socially inappropriate when an out-group member was involved.

We measured social appropriateness using the incentivized elicitation method developed by Krupka and Weber (2013). A group of 53 subjects was asked to evaluate the appropriateness of the action "roll after you made a promise" or "do not roll after you did not make a promise" in different social-identity contexts. Possible answers were five: very socially inappropriate, somewhat socially inappropriate, neutral, somewhat socially appropriate, very socially appropriate. After rating all actions in all choice scenarios, subjects were paid privately. They received a participation fee and an additional payment related to their evaluation. They received the additional payment if their appropriateness rating matched the mode in one of the experimenter's randomly selected two possible actions.⁴³

Defining the possible action by a_k with $k \in \{1, 2\}$, we assume that a participant cares about the monetary payoff produced by the selected action, $\pi(a_k)$, and the degree to which the action is collectively perceived as socially appropriate, $N(a_k)$. As in Krupka-Weber (2013), we assume that monetary payoffs and social norms linearly affect the dictator's utility:

$$u(a_k) = \beta\pi(a_k) + \gamma N(a_k)$$

where $\beta \geq 0$ is the value the participant places on the monetary payoff, and $\gamma \geq 0$ captures the degree to which the participant cares about adhering to social norms.

³⁹ In particular, a red promisor dictator is likely to keep her promises and, when switched, those made by red dictators to red or blue recipients and those made by red or blue dictators to a red recipient. However, she is less likely to keep a promise made by blue dictators to blue recipients.

⁴⁰ It is worth noting that this result cannot be used to investigate motivations for promise-keeping since it can be driven by two potential differences in our comparisons (i.e., the dictator's non-switch/switch condition and the switched recipient's in-group/out-group condition, which can affect his second-order beliefs.)

⁴¹ We thank a referee for suggesting us this extension.

⁴² Social norms generally prescribe behaviors or actions rather than outcomes. They must be jointly recognized or collectively perceived by members of a population. Social norms prescribe what one "should do" or "should not do," Krupka and Weber (2013) thus define them as collective perceptions among members of a population regarding the appropriateness of different behaviors.

⁴³ See Krupka and Weber (2013: Experiment 1) for details. We strictly adapted their method to our case.

Table 5
Relative social appropriateness.

	Socially inappropriate		Socially Neutral	Socially appropriate		Appropriateness Rating
	Very	Somehow		Somehow	Very	
Roll in-group	0.0 %	7.6 %	1.9 %	47.2 %	43.4 %	0.63
Don't Roll in-group	18.9 %	56.6 %	9.4 %	13.2 %	1.9 %	−0.39
Roll out-group	1.9 %	3.8 %	5.7 %	34.0 %	54.7 %	0.68
Don't Roll out-group	0.0 %	41.5 %	5.7 %	15.1 %	37.7 %	0.25

Note: The variable “appropriateness rating” converts subject survey responses to numerical scores (“very socially inappropriate” = −1, “somewhat socially inappropriate” = −1/2, “neutral” = 0, “somewhat socially appropriate” = 1/2, “very socially appropriate” = 1.).

We adopt survey data to measure the perceived appropriateness rating (last column of Table 6) and use our experimental-game outcomes to estimate the two weights β and γ . Given the linear restrictions, we use conditional logit. We report the results in Table 6.⁴⁴

The dependent variable is the decision on the action a_k , i.e., if the action is chosen (1) or not (0.) As stated, the possible actions for each subject are “roll after you made a promise” or “do not roll after you made a promise.” In columns (1) and (2), we consider two specifications with and without an interaction effect between identity and social norms.

The positive coefficient β indicates that people care about own monetary payoff, while the positive coefficient γ implies that subjects consider the appropriateness of norms in making their decisions. The impact of social norms, however, depends on the situation considered. Specifically, the interaction between the appropriateness ratings and social identity (δ) is statistically significant. Furthermore, in making their choices, experimental subjects care about the social appropriateness elicited from our survey, doing so differently when in-group or out-group. Finally, we compute the sacrifice ratio, which identifies how many tokens an individual is willing to sacrifice to take a “very socially appropriate action” rather than a “very socially inappropriate one.” In column (2), this ratio shows that subjects are willing to pay 0.58 tokens to comply with social norms but that this increases to 2.10 tokens when the subject interacts with a group member, as stated before.⁴⁵

Results from the interaction between observed choices and survey data show that people react to the perception of the social appropriateness of their choice and that their relevance is differently perceived when they are in-group or out-group. We can summarize as follows.

Result 6. People care more about following the perceived social norm when paired with an in-group member.

6. Conclusion

In this paper, we constructed a weak mechanism of group favoritism (color assignment, a form of the minimal-group paradigm) and augmented it with communication. Since pairs were switched half the time, this initial communication is direct (with no switch) or indirect (with a switch).

We test for differences in-group versus out-group allocations when there is no direct communication and compare these differences to those found with direct communication. In both cases, we find considerable support for in-group favoritism. Dictators in-group matches are more likely to roll than dictators in out-group matches. In addition, experimental outcomes support the idea that direct communication leads to more people being more likely to sacrifice own payoffs in favor of the co-player. However, they do not support the notion that in-group favoritism leads to different effects when people directly communicate with each other, as, in this circumstance, the effect sizes for in-group and out-group cases are similar.

We then focused on communication content, specifically mutual promises (or agreements) and other messages. There is no difference in the rates of mutual promises made with in-group versus out-group members. However, we confirm favoritism towards one's color group when there are no mutual promises to act favorably: Dictators are 11 percentage points more likely to sacrifice own payoffs when they are paired with same-color recipients than with different-color recipients, so our relatively modest identity inducement was effective. In principle, there could be a null effect of communication on this gap or even a negative one if communication tended to exacerbate differences or strengthen in-group ties. Mutual promises are highly effective, with favorable dictator choices made nearly 70 % of the time.

The significant difference across recipient color matches completely disappears when we consider only those cases with mutual promises to *Roll*. After direct communication, subjects are, in fact, likely to keep their own promises independently of their partners' color. Out-group and in-group promisors, however, appear to be motivated by different mechanisms. We test moral commitment as a motivation for promise-keeping. Our results prove that out-group promisors' behavior is consistent with a moral commitment explanation. By contrast, the behavior of in-group subjects is not but is consistent with expectation-based explanations, such as one based on a sense of guilt.

⁴⁴ We bootstrap standard errors for the coefficients to account for our estimates of appropriateness being noisy.

⁴⁵ The sacrifice ratio is derived by computing the difference between the two extreme cases for social appropriateness: “very socially appropriate action” and “very socially inappropriate one,” and reporting it in terms of payoffs by using β . Formally, from the coefficient of column (2), we compute $1 * \gamma - (-1 * \gamma) / \beta = 2\gamma / \beta = 0.58$ for out-groups and $1 * (\gamma + \delta) - (-1 * (\gamma + \delta)) / \beta = 2(\gamma + \delta) / \beta = 2.10$ for in-groups. We test that the ratio is significantly different from zero. Results are available upon request.

Table 6
Subject's choices: conditional logit estimation with mean appropriateness ratings.

	(1)	(2)
Monetary payoff (β)	0.728*** (0.159)	0.642*** (0.115)
Appropriateness rating (γ)	0.780* (0.411)	.186** (0.088)
Identity \times Appropriateness (δ)		.491** (0.232)
Log-likelihood	−310.05	−309.29
Obs. (subjects)	792 (246)	792 (246)

Note: (*) $p < 0.10$; (**) $p < 0.05$; (***) $p < 0.01$; all two-tailed. Bootstrapped standard errors are in parentheses.

Nevertheless, exploring experimental outcomes, the interaction between identity and communication seems more complex than we can determine with our experimental design.⁴⁶ We mention the notion of transforming out-group members to being perceived as in-group members. Does this reflect a reduction in the social distance due to the color category, or does it reflect a similarity in a different identity (promise-keepers)?

So, mutual promises not only lead to more willingness to make pro-social choices in all pairs (whether switched or not) but also reduce or eliminate in-group favoritism. Since, in this case, there were no differences in roll rates between in-group and out-group subjects, everyone was treated the same regardless of color. We consider this to be a hopeful sign. Having direct communication here bridged a gap. Perhaps this finding or something similar would also apply in field environments. We invite others to investigate further the rich interaction between communication and social identity.

Declaration of Competing Interest

None.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.geb.2023.12.005](https://doi.org/10.1016/j.geb.2023.12.005).

References

- Adnan, W., Arin, K.P., Charness, G., Lacomba, J.A., Lagos, F., 2022. Which social categories matter to people: an experiment. *J. Econ. Behav. Organ.* 189, 125–145.
- Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. *Q. J. Econ.* 115, 715–753.
- Akerlof, G., Kranton, R., 2002. Identity and schooling: Some lessons for the economics of education. *J. Econ. Lit.* 40 (4), 1167–1201.
- Akerlof, G., Kranton, R., 2005. Identity and the economics of organizations. *J. Econ. Perspect.* 19 (1), 9–32.
- Bénabou, R., Tirole, J., 2011. Identity, morals, and taboos: beliefs as assets. *Q. J. Econ.* 126 (2), 805–855.
- Ben-Ner, A., McCall, B.P., Stephane, M., Wang, H., 2009. Identity and in-group/out-group differentiation in work and giving behaviors: experimental evidence. *J. Econ. Behav. Organ.* 72, 153–170.
- Brandts, J., Charness, G., Ellman, M., 2016. Let's talk: how communication affects contract design. *J. Eur. Econ. Assoc.* 14, 943–974.
- Brewer, M.B., 1979. In-group bias in the minimal group situation: a cognitive-motivational analysis. *Psychol. Bull.* 86, 307–324.
- Bourhis, R.Y., Gagnon, A., Brown, R., Gaertner, S., 2001. Social orientations in the minimal group paradigm. *Intergroup Processes*, Blackwell Handbook in Social Psychology, vol. 4. Oxford, Blackwell, pp. 89–111.
- Cason, T.N., Lau, S.H.P., Mui, V.L., 2019. Prior interaction, identity, and cooperation in the inter-group Prisoner's Dilemma. *J. Econ. Behav. Organ.* 166, 613–629.
- Charness, G., 2000. Self-serving cheap talk: a test of Aumann's conjecture. *Games Econ. Behav.* 33, 177–194.
- Charness, G., Dufwenberg, M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.
- Charness, G., Dufwenberg, M., 2010. Bare promises: an experiment. *Econ. Lett.* 10, 281–283.
- Charness, G., Villeval, M., 2009. Cooperation and competition in intergenerational experiments in the field and the laboratory. *Am. Econ. Rev.* 99, 956–978.
- Charness, G., Holder, P., 2018. Charity in the laboratory: matching, competition, and group identity. *Manag. Sci.* 65, 1398–1407.

⁴⁶ Exploratory evidence based on our experimental data, e.g., shows that subjects tend to keep promises made or received by others of their in-group members. By contrast, when someone makes promises from a different group, subjects only support recipients belonging to their group, while they do not support those belonging to a different group. Experimental outcomes are thus consistent with the idea that subjects are more likely to feel guilt aversion when facing someone belonging to their group or feeling somehow committed to keeping promises by others only when they belong to their group.

- Charness, G., Chen, Y., 2020. Social identity, group behavior, and teams. *Annu. Rev. Econom.* 12, 691–713.
- Charness, G., Feri, F., Meléndez-Jiménez, M.A., Sutter, M., 2021. An experimental study on the effects of communication, credibility, and clustering in network games. forthcoming in *Rev. Econ. Stat.*
- Charness, G., Rigotti, L., Rustichini, A., 2007. Individual behavior and group membership. *Am. Econ. Rev.* 97, 1340–1352.
- Charness, G., Cobo-Reyes, R., Jiménez, N., 2014. Identities, selection, and contributions in a public-goods game. *Games Econ. Behav.* 87, 322–338.
- Chen, Y., Li, S.X., 2009. Group identity and social preferences. *Am. Econ. Rev.* 99, 431–457.
- Chen, Y., Li, S.X., Liu, T., Shih, M., 2014. Which hat to wear? Impact of natural identities on coordination and cooperation. *Games Econ. Behav.* 84, 58–86.
- Ciccarone, G., Di Bartolomeo, G., Papa, S., 2020. The rationale of in-group favoritism: an experimental test of three explanations. *Games Econ. Behav.* 124, 554–568.
- Cooper, R., DeJong, D.V., Forsythe, R., Ross, T.W., 1989. Communication in the battle of the sexes game: some experimental results. *RAND J. Econ.* 20, 568–587.
- Daskalova, V., 2018. Discrimination, social identity, and coordination: an experiment. *Games Econ. Behav.* 107, 238–252.
- Deutsch, M., 1958. Trust and suspicion. *J. Confl. Resolut.* 2 (4), 265–279.
- Di Bartolomeo, G., Dufwenberg, M., Papa, S., 2019a. The sound of silence: a license to be selfish. *Econ. Lett.* 182, 68–70.
- Di Bartolomeo, G., Dufwenberg, M., Papa, S., Passarelli, F., 2019b. Promise, expectations and causation. *Games Econ. Behav.* 113, 137–146.
- Di Bartolomeo, G., Dufwenberg, M., Papa, S., Passarelli, F., 2023a. Promises or agreements? Moral commitments in bilateral communication. *Econ. Lett.* 222 (1).
- Di Bartolomeo, G., Dufwenberg, M., Papa, S., 2023b. Promises and partner-switch. *J. Econ. Sci. Assoc.* 9, 77–89.
- Ellingsen, T., Johannesson, M., 2004. Promises, threats and fairness. *Econ. J.* 114, 397–420.
- Feri, F., Irlenbusch, B., Sutter, M., 2010. Efficiency gains from team-based coordination: Large-scale experimental evidence. *American Economic Review* 100 (4), 1892–1912.
- Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178.
- Güth, W., Ploner, M., Regner, T., 2009. Determinants of in-group bias: is group affiliation mediated by guilt-aversion? *J. Econ. Psychol.* 30, 814–827.
- Hargreaves Heap, S.P., Zizzo, D.J., 2009. The value of groups. *Am. Econ. Rev.* 99, 295–323.
- Ioannou, C.A., Qi, S., Rustichini, A., 2015. Group payoffs as public signals. *J. Econ. Psychol.* 48, 89–105.
- Klor, E.F., Shayo, M., 2010. Social identity and preferences over redistribution. *J. Public Econ.* 94, 269–278.
- Kranton, R., Sanders, S., 2017. Groupy vs. non groupy social preferences: personality, region, and political party. *Am. Econ. Rev. Pap. Proc.* 107, 65–69.
- Kocher, M.G., Praxmarer, M., Sutter, M., 2020. Team decision-making. In: Zimmermann, K. (Ed.), *Handbook of Labor, Human Resources and Population Economics*. Springer, pp. 1–25.
- Kranton, R., Pease S. Sanders, M., Huettel, S., 2018. Groupy vs. Non-Groupy behavior: Deconstructing group Bias. Duke University, mimeo.
- Krupka, E.L., Leider, S., Jiang, M., 2017. A meeting of the minds: informal agreements and social norms. *Manag. Sci.* 63 (6), 1708–1729.
- Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? *J. Eur. Econ. Assoc.* 11 (3), 495–524.
- Liebe, U., Tutic, A., 2010. Status groups and altruistic behaviour in dictator games. *Ration. Soc.* 22, 353–380.
- Loomis, J.L., 1957. Communication and the Development of Trust. New York University. Doctoral dissertation.
- Loomis, J.L., 1959. Communication, the development of trust, and cooperative behavior. *human relations. Hum. Relat.* 12 (4), 305–315.
- Ockenfels, A., Werner, P., 2014. Beliefs and ingroup favoritism. *J. Econ. Behav. Organ.* 108, 453–462.
- Pan, X.S., Houser, D., 2013. Cooperation during culture group formation promotes trust towards members of out-groups. *Proc. R. Soc. B Biol. Sci.* 280 (1762).
- Pan, X.S., Houser, D., 2019. Why trust out-groups? The role of punishment under uncertainty. *J. Econ. Behav. Organ.* 158, 236–254.
- Shayo, M., 2009. A model of social identity with an application to political economy: nation, class, and redistribution. *Am. Polit. Sci. Rev.* 103 (2).
- Shih, M., Pittinsky, T.L., Ambady, N., 1999. Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychol. Sci.* 10, 80–83.
- Tajfel, H., 1970. Experiments in intergroup discrimination. *Sci. Am.* 223, 96–102.
- Tajfel, H., 1970. *Social Identity and Intergroup Relations*. Cambridge University Press, New York, USA.
- Tajfel, H., Turner, J.C., 1979. An integrative theory of intergroup conflict. In: Austin, W.G., Worchel, S. (Eds.), *The Social Psychology of Intergroup Relations*. Brooks/Cole, Monterey, pp. 33–47.
- Tajfel, H., Turner, J.C., 1986. The social identity theory of intergroup behavior. In: Worchel, S., Austin, W. (Eds.), *Psychology of Intergroup Relations*. Nelson Hall, Chicago, pp. 7–24.
- Tajfel, H., Billig, M., 1973. Social categorization and similarity in intergroup behaviour. *Eur. J. Soc. Psychol.* 3, 27–52.
- Tajfel, H., Turner, J.C., Austin, W.G., Worchel, S., 1979. An integrative theory of intergroup conflict. In: Austin, W.G., Worchel, S. (Eds.), *The Social Psychology of Intergroup Relations*. Brooks/Cole, Monterey, pp. 33–47.
- Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C., 1971. Social categorization and intergroup behavior. *Eur. J. Soc. Psychol.* 1, 149–178.
- Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76, 1467–1480.
- Whitt, S., Wilson, R.K., 2007. The dictator game, fairness and ethnicity in postwar Bosnia. *Am. J. Pol. Sci.* 51, 655–668.