

Dhami, Sanjit; Wei, Mengxing; Mamidi, Pavan

Working Paper

Religious Identity, Trust, Reciprocity, and Prosociality: Theory and Evidence

CESifo Working Paper, No. 10147

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dhami, Sanjit; Wei, Mengxing; Mamidi, Pavan (2022) : Religious Identity, Trust, Reciprocity, and Prosociality: Theory and Evidence, CESifo Working Paper, No. 10147, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/271791>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Religious Identity, Trust, Reciprocity, and Prosociality: Theory and Evidence

Sanjit Dhami, Mengxing Wei, Pavan Mamidi

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Religious Identity, Trust, Reciprocity, and Prosociality: Theory and Evidence

Abstract

We use the trust and the dictator games to explore the effects of religious identity on trust, trustworthiness, prosociality, and conditional reciprocity within a beliefs-based model. We provide a novel and rigorous theoretical model to derive the relevant predictions, which are then tested in pre-registered lab-in-the-field experiments from villages in the Indian states of Bihar and Uttar Pradesh. We find strong evidence of religious identity effects in the beliefs, and the chosen actions, for Hindu and Muslim subjects. Priming has little effect on Hindu subjects but it enhances religious polarization in beliefs and actions among Muslim subjects. There is taste-based discrimination but no statistical discrimination. All our underlying assumptions on beliefs, and their dependence on priming and identity are confirmed by the data, identifying a precise beliefs-based mechanism for the effects of religious identity. More religious subjects expect greater prosociality/reciprocity and often are more prosocial/reciprocal.

JEL-Codes: C910, D010, D840, D910.

Keywords: religious identity, trust, trustworthiness, prosociality, priming, conditional reciprocity.

*Sanjit Dhami**

*Division of Economics, School of Business
University of Leicester, London Road
United Kingdom – Leicester, LE2 1RQ
sd106@leicester.ac.uk*

Mengxing Wei

*School of Economics, The Laboratory for
Economic Behaviors and Policy Simulation
Nankai University / Tianjin / China
mengxing.wei@hotmail.com*

Pavan Mamidi

*Centre for Social and Behavior Change,
CSBC (Ashoka University)
India – New Delhi, Delhi 110057
pavan.mamidi@ashoka.edu.in*

*corresponding author

December 7, 2022

1 Introduction

There has been widely noted growth in right wing political parties around the world, characterized by their radicalism/extremism, populism, and nationalism (Golder, 2016). Right wing populism, a term that combines right-wing politics and an appeal to populist policies and rhetoric is now increasingly used to describe political developments in Europe, Latin America, Africa, and parts of Asia, such as in the Middle-East and South-Asia. Political parties that have these features are often anti-immigrant, Islamophobic, ethno-nationalist, and thrive on real or perceived threats to national identity and economy from outsiders.

India, a democracy, which houses one-sixth of the world population, is an important case study of right wing populism. In the last decade there have been allegations of a desire to shape economic and social policy on the basis of a particular interpretation of national identity. These allegations relate to the exclusion or marginalization of religious minorities at the expense of a pan-India Hindu-based identity; pressure on secular institutions; increasing difficulty in defending the idea of a secular and inclusive India for opposition parties; and growing incidents of hatred and violence against minority communities, particularly the Muslims (Vaishnav, 2019; Sahoo, 2020; Pollard, 2022).

Social scientists and the international press have been active in describing and understanding contemporary Indian polity and society. However, there is a paucity of economic analyses of various aspects of religious polarization in India; we aim to address this gap. We explore the effects of religious identity on trust, trustworthiness, conditional reciprocity, and prosociality, in India. We use an explicitly beliefs-based model to provide the microfoundations for the observed actions of the subjects. Our model is motivated by *psychological game theory*, hence, it requires us to carefully outline the beliefs players have about the actions of others (first order beliefs), and beliefs about the first order beliefs of others (second order beliefs).¹ We use a lab-in-the-field study with 542 Hindu and Muslim subjects drawn from villages in Bihar and Uttar Pradesh in India.² We conduct a trust game and a dictator game with our subjects. Our control treatment group is unprimed and our treatment group is primed for a religious identity. We also distinguish between *taste based discrimination* and *statistical discrimination*.

1.1 Motivation for the paper

The existing literature is entirely empirical. No formal rigorous beliefs-based theoretical model has been proposed that takes simultaneous account of trust, trustworthiness, conditional reciprocity, prosociality, identity, and priming. We propose the relevant economic theory in this paper, which is an advance in the literature on identity economics, in order to discipline and

¹A distinguishing feature of such models, that play a central role in modern behavioral game theory, is that beliefs of various orders directly enter into the utility functions of players and determine their actions; see Geanakoplos et al. (1989), Rabin (1993), Dufwenberg and Kirchsteiger (2004), Battigalli and Dufwenberg (2009). For a recent survey, see Battigalli and Dufwenberg (2022), and for textbook treatment, see Dhami (2020; Vol IV).

²A purely field study, based on observational data, is unlikely to be able to successfully engage in the ‘belief manipulation’ that is required for testing the predictions of our model. This requires the direct measurement of first order and second order beliefs of subjects and testing how these beliefs are directly influenced by priming and social identity.

organize the empirical exercise that follows it. We take explicit account of the beliefs of the players; how these beliefs are influenced by religious identity and priming; and how beliefs, in turn, determine trust, trustworthiness, conditional reciprocity, and prosociality. The model predicts that beliefs, conditional reciprocity, religious identity, and priming, interact and determine actions. Our lab-in-the-field study in India, provides a stringent empirical test of our model, and supports our predictions and assumptions. To the best of our knowledge, this is the first direct measurement and confirmation of a beliefs-based mechanism in influencing the effects of identity, despite the voluminous literature on the subject; for a survey, see Dhami (2019, Vol. II).

The two-player trust game, due to Berg et al. (1995), is the standard tool used to measure trust and trustworthiness. The amount of investment sent by the trustor (player 1) is a measure of *trust* and the amount returned by the trustee (player 2) is a measure of *trustworthiness*. Trust is essential for efficient economic outcomes (Arrow, 1972). Greater trust may be beneficial for cooperation, improved economic and social outcomes, growth, and political stability (Fukuyama, 1995; Knack and Keefer, 1997; Easterly and Levine, 1997; Zak and Knack, 2001; Algan and Cahuc, 2014). Alesina and LaFerrara (2002) show that minorities who have been discriminated against, trust less, a finding borne out by several subsequent studies. The dictator game is an experimental game, designed to measure the extent of altruism towards a passive receiver. For surveys of the results from both games, see Dhami (2019, Vol. II).

Religious affiliation is our measure of identity.³ It might also capture the effect of other underlying variables such as childhood home environment, or deep underlying preferences and tastes. Hence, in order to establish causation between religious affiliation and economic behavior, one needs to create an exogenous variation in religion, and control for other variables. This is typically achieved by having a control group and a treatment/primed group to which subjects are randomly assigned (Shariff et al., 2016; Benjamin et al., 2016).

There are several alternative methods of measuring religiosity (Benjamin et al., 2016). This includes unidimensional measures, e.g., how often does the subject pray or visit a temple to offer prayers (Gupta et al. 2018). The meta study by Shariff et al. (2016) finds weak effects of religion on prosocial behavior.⁴ One potential explanation is that religion is essentially a multidimensional variable and using unidimensional measures is unreasonable.⁵ We use a multidimensional religiosity measure, due to Rohrbaugh and Jessor (1975), that incorporates four aspects of religiosity - ritual, consequential, ideological, and experiential. In differentiating between religious ingroups and outgroups, we directly reveal the religion of the ‘other player’ in a trust game or a dictator game as in Chuah et al. (2014), and Gupta et al. (2018).⁶

³A potential worry might be that more prosocial/trusting/trustworthy people might self-select themselves into a particular religion. However, people rarely change their religions, so this has little empirical relevance.

⁴The meta study by Lane (2016) also finds mixed results about ingroup/outgroup effects of religion. There are relatively small effects of religious outgroup discrimination in 14.3% of the studies, no effects in 80.9% studies, and outgroup favoritism is found in 4.8% studies.

⁵Using a multidimensional measure, Tan and Vogel (2008) found that more religious German University students were trusted more. A multidimensional measure also reveals greater cooperation in a prisoner’s dilemma game among Malaysian student subjects (Chuah et al., 2014).

⁶In some papers, the names of ‘other players’ that reflect their religion are revealed (Chakravarty et al., 2016; Chuah et al., 2013; Fershtman and Gneezy, 2001). One drawback of subject names is that they ignore the religious

Observed differences in trust and trustworthiness towards members of a particular religion do not necessarily reveal that there is *taste based discrimination*, i.e., underlying prejudices that are part of preferences (Becker 1957). Rather, these differences might reveal *statistical discrimination*, which arises when the social/religious/ethnic affiliation of a player gives valuable signals about their actual behavior (Arrow, 1973, 1998). Unlike the findings in Fershtman and Gneezy (2001) and Chuah et al. (2016), we do not find any evidence of statistical discrimination in India.

1.2 Findings

Hindu and Muslim subjects respond very differently to priming, in terms of their beliefs and their actions.⁷ Priming has little effect on the beliefs and actions of Hindu subjects. However, priming sharpens the ingroup/outgroup conflict in beliefs and in actions for Muslim subjects, in their roles as trustors, trustees, and dictators in the two games. We organize our findings under several headings.

1. *First order beliefs* (FOB): The FOB of trustors are their expectations about the amount to be returned by the trustee, i.e., their beliefs about trustworthiness. In the neutral treatment, 67% of the Muslim trustors had identical FOB's about the trustworthiness of ingroup trustees (Muslims) and outgroup trustees (Hindus). This reduces to 30% in the primed treatment; a statistically significant reduction. In the neutral treatment, 25% of Muslim trustors assign strictly greater trustworthiness to ingroup trustees and this increases to 55% in the primed treatment. By contrast, whether primed or unprimed, Hindu trustors believe that 57% of ingroup trustees (Hindus) will return a strictly higher amount than outgroup trustees (Muslims).
2. *Second order beliefs* (SOB): The SOB of the trustees are their beliefs about what return the trustor expects from them, i.e., the beliefs of the trustees, about the trustworthiness expected from them by trustors. Priming Muslim trustees significantly increases their SOB, i.e., their expectations of how much ingroup trustors expect them to return, relative to the neutral treatment; thereby increasing their expectations of religious polarization. However, for Hindu trustees, priming reduces expectations of such ingroup/outgroup polarization, when measured by their SOB.⁸

sub-groups within religions. For instance, the names might reveal if a Hindu is a Brahmin or a Shudra, or a Muslim is a Shia or a Sunni. Given the traditional divides between these sub-groups and potential discrimination across subgroups in an Indian context, giving out names may add extra noise to the results.

⁷Priming took the form of providing factual and readily publicly available information on communal riots in India. We do not offer an interpretation of that data to our subjects.

⁸Our data does not allow us to pin down the reason for the differences in the responses of SOB of Muslim and Hindu subjects. However, we may offer a conjecture. In many riots in India, several studies indicate that the Muslims disproportionately face communal violence in India with the active connivance of the state machinery. See, for instance, the Guardian article "Inside Delhi: beaten, lynched and burnt alive" dated 1 March 2020, describing one of the most recent incidents of communal riots in India. As such, priming with information about riots may make salient, the unfair treatment of Muslims. Since Muslims are typically the aggrieved party, they are likely to become more polarized with priming. By contrast, many Hindu subjects might feel empathy with the treatment of Muslims, so priming them with information on riots induces an empathy-response, potentially reducing their degree of religious polarization towards Muslims. Testing this precise mechanism, which was not an objective of our experimental design, will require a separate study.

3. *Trustees*: The percentage of Hindu trustees who return relatively more to ingroup trustors in the primed and the neutral treatments is nearly identical, 66% and 65%. By contrast, only 5% of the Muslim trustees in the neutral treatment returned relatively more to the ingroup trustors, but this increases massively to 59% in the primed treatment. A significant majority, 63%, of the Muslim trustees in the neutral treatment returned exactly the same amounts to ingroup and outgroup trustors, but this falls to 18% in the primed treatment. We also show that for both Hindu and Muslim subjects, the data confirms the *marginal identity effects*, which are predicted by our theoretical model.⁹
4. *Trustors*: More than 85% of the trustors send the same, or higher, amounts to ingroups. 57% of the Hindu trustors send strictly greater investment to ingroup trustees in the neutral treatment, and this falls only slightly to 55% when they are primed. The percentage of Hindu trustors who send identical amounts to ingroup and output trustees is also statistically indistinguishable in the primed and the neutral treatments (32% and 29%). However, in the neutral treatment, only 17% Muslim trustors send more investment to their ingroups but this figure increases significantly to 59% in the primed treatment. In the neutral treatment, 79% of Muslim trustors send identical amounts to ingroup and outgroup trustees, but this figure falls significantly to 32% when primed.
5. *Social identity concerns*: Both Hindu and Muslim trustors invest more to ingroup members than outgroup members. This is mirrored in the return behavior of trustees. However, the average investments sent to ingroup (and outgroup) trustees is similar for Hindu and Muslim trustors. Similarly, the average amounts returned to ingroup (and outgroup) trustors is similar for Hindu and Muslim trustees. Thus, there is also a common pattern of behavior among Hindu and Muslim subjects, possibly arising due to shared cultural and social factors.
6. *Dictator game*: In the dictator game, priming has a statistically insignificant effect on the ingroup/outgroup transfers of Hindu dictators. However, priming sharpens the difference between ingroup/outgroup transfers of Muslim dictators. As noted above, this pattern is also observed for the trust game. The portability of beliefs and actions across very different games appears to capture deep underlying preferences.
7. *Taste-based or statistical discrimination*: We establish that for both Hindu and Muslim subjects, there is no statistical discrimination in trust or in trustworthiness, and all observed discrimination (ingroup/outgroup differences) is taste based.
8. *Direct confirmation of assumptions on beliefs*: Our theoretical model rests on three key assumptions on beliefs (see Section 2). We believe that we are the first to confirm these assumptions by direct analysis of the beliefs data. The more sophisticated econometric analysis (see Section 11) also largely confirms these assumptions. This lays a firm beliefs-based foundation for observed differences in ingroup/outgroup behavior.

⁹We have several kinds of marginal identity effect that are defined more precisely in our formal model. For trustees, this is the difference in the amount returned to ‘ingroup’ trustors when primed and when not primed, minus the difference in amounts returned to ‘outgroup’ trustors when primed and when not primed.

9. *Results from regression analysis:* We conduct a regression analysis of the determinants of beliefs and actions in the trust game separately (on FOB, SOB, trust, and trustworthiness), in addition to an analysis of the determinants of dictator transfers. We confirm once again the assumptions on our beliefs, and ingroup/outgroup differences in behavior, after controlling for a range of other variables. We also demonstrate the existence of various kinds of marginal identity effects.¹⁰ We omit a discussion of these effects below and focus on the other determinants.

SOB of trustees: The SOB of trustees reveal that they believe that trustors expect a higher return if trustors choose higher investment. Thus, they anticipate expectations of reciprocity among trustors; such anticipation is relatively greater among Muslim trustees. Trustees with higher scores on the multidimensional religiosity measure believe that trustors expect greater trustworthiness; this behavior is more pronounced among Muslim trustees. More educated Muslim trustees have higher SOB, but education has no effect on the SOB of Hindu trustees. There are also significant differences in the SOB of Hindus and Muslims trustees with respect to age.

Amounts returned by trustees: Both Hindu and Muslim trustees exhibit conditional reciprocity in the amounts returned to trustors (i.e., the higher is the investment, the higher is the amount returned); Muslim trustees are slightly more reciprocal. More religious trustees return more; the effects are stronger for Hindu trustees. Older, female, unmarried, less educated trustees also return higher amounts to the trustors; but the significance of these variables differs for Hindu and Muslim trustees.

FOB of trustors: Trustors anticipate the conditional reciprocity of trustees because their FOB reveal that trustees are expected to return more if trustors choose higher investment; Muslim trustors expect greater reciprocity. Trustors who are more religious expect more trustworthiness from trustees; but this effect is only significant for Muslim trustors. Female Hindu trustors expect lower returns than males but gender has no effect on the FOB of Muslim trustors; a similar qualitative difference is observed for marital status. More educated Hindu trustors expect higher return; but education plays no role in the FOB of Muslim trustors.

Amounts invested by trustors: A probit analysis shows that more educated Hindu trustors are likely to send lower investment to the trustees; but education does not explain the behavior of Muslim trustors. Religiosity, Gender, marital status, and age, are not significant factors in explaining the behavior of trustors. However, we confirm the ingroup/outgroup differences in behavior, and the expectations of reciprocity on the part of trustors.

Dictator game: Very few variables, other than the classical religious social identity effects (i.e., ingroup/outgroup differences) are significant in the analysis of dictator giving. Hindu dictators transfer more than Muslim dictators. Married Hindu dictators transfer relatively more to the receivers; but marital status is insignificant for Muslim dictators.

¹⁰As noted above, this refers to differences in beliefs and actions towards ingroups/outgroups when one is primed, relative to when one is not primed.

1.3 Relation to the existing literature

For a useful survey of the existing literature that measures the effect of religious identity on observed choices see Benjamin et al. (2016).¹¹ Shariff and Norenzayan (2007) and Ahmed and Salas (2011) find that priming for religion increases generosity in a dictator game. In a dictator game with students in Granada, Brañas-Garza et al. (2014) find that religiously active Catholics are relatively more generous, and Catholics have higher minimum acceptable offers in the ultimatum game as compared to any of the other categories.¹² Adida et al. (2016) show that second and third generation French immigrant subjects discriminate in a trust game against Senegalese Muslim immigrants relative to Christian immigrants. For Chinese data there is evidence of ingroup/outgroup discrimination in a trust game (Mantilla et al., 2021) and also differential effects of such discrimination among members of different religions (Xia et al., 2021).

Several studies use data from the Indian subcontinent, but they do not use religious priming, and the analysis is entirely empirical. Johansson-Stenman et al. (2009) find no effect of religious identity in a trust game among Bangladeshi Hindu and Muslim subjects, where ethnic conflict is low. Chuah et al. (2013) conduct a trust game with 129 Muslim and Hindu subjects in Mumbai and find significant ingroup/outgroup effects. Gupta et al. (2018) show, in a trust game and a dictator game in Bangladesh and Bengal in India, that ‘relative status’ (whether one is in a majority or minority) plays an important role. Chakravarty et al. (2016) play the prisoner’s dilemma and stag hunt games in villages in Bengal and show that cooperation rates depend on the extent of religious fragmentation in the villages.

1.4 Organization of the paper

In Section 2, we describe the trust game and the preferences and beliefs of the players. Section 3 derives an expression for the conditional reciprocity of the trustees. Sections 4 and 15.1 give, respectively, the solutions to the trust game and the dictator game, and derive the theoretical predictions. Section 6 describes the experimental design. Section 7 gives direct empirical tests of the three key assumptions on beliefs in Section 2. Section 8 tests the predicted effects of social identity and priming on trust and trustworthiness that were given in Section 4, while Section 9 tests similar predicted effects in the dictator game, given in Section 15.1. Section 10 tests for taste based and statistical discrimination. Section 11 takes the predicted determinants of actions and beliefs in the trust game in Section 4, and subjects them to a formal econometric analysis. Section 12 repeats this exercise for the dictator game for the predicted determinants in Section 15.1. The final section concludes. All proofs are in the appendix.

¹¹Some of the existing research does not find a significant effect of religiosity on behavior (Fehr et al., 2003; Karlan, 2005; Anderson and Mellor, 2009; Anderson et al., 2010).

¹²Putnam (1993) argued that the Catholic religion is more centralized and organized in vertical hierarchical terms, so it reduces trust, while Protestant churches are more autonomous and horizontally organized, which increases trust. Previous empirical literature shows that trust and public goods contributions among Catholics are lower than Protestants (Putnam, 1993; Alesina and La Ferrara, 2002; Guiso, Sapienza, and Zingales, 2003; Arruñada, 2010). Trauttmüller (2011) finds that for German data, Protestants trust more as compared to Catholics, and people who attend religious services also trust more. People who are actively integrated into religious networks also trust more. Benjamin et al. (2016) find that trust among Protestants is not influenced by priming. Priming causes Protestants to increase contributions to public goods, whereas Catholics decrease their contributions.

2 Preferences and beliefs in the trust game

Consider a two-player trust game, augmented to include the roles of religious identity and priming. For any player, the identity of the other player is indexed by $s \in \{0, 1\}$, where $s = 1$ for an ingroup member and $s = 0$ for an outgroup member. For instance, for a Hindu trustor, a Hindu trustee is an ingroup member ($s = 1$) and a Muslim trustee is an outgroup member ($s = 0$). Define the binary priming variable $p \in \{0, 1\}$ such that $p = 1$ for subjects ‘primed’ for a religious identity, and $p = 0$ for unprimed subjects.

2.1 Preferences

Player 1, the trustor, has an experimenter-provided endowment $Y > 0$. The trustor can send any part of the endowment as an *investment*, $i \in [0, Y]$, to Player 2, the trustee, who has zero endowment. The experimenter triples the investment before passing it on to the trustee. Hence, the trustee receives $3i \geq 0$. The trustee can now *return* any amount, $r \in [0, 3i]$, to the trustor. The amount of investment, i , is a measure of *trust* and the amount returned, r , is a measure of *trustworthiness*. The material payoff of the trustor is $Y - i + r$, and the material payoff of the trustee is $3i - r$. When the trustee is passive, so by default $r = 0$, we get a dictator game, as in Fershtman and Gneezy (2001). We consider the predictions of this game in Section 15.1.

The trustor has self-regarding preferences from material payoffs, $Y - i + r$. The trustor’s utility function is¹³

$$U(i, r) = Y - i + r. \quad (2.1)$$

The utility function of the trustee is given by

$$V(i, r) = v(3i - r) + \lambda R; \lambda > 0, v' > 0, v'' \leq 0. \quad (2.2)$$

The trustee derives utility from *material payoffs*, $3i - r$, and from *conditional sequential reciprocity*, R , after having observed the trustor’s investment, i ; $\lambda > 0$ is the relative weight assigned to reciprocity. A formalization of R requires specifying beliefs of various orders (Section 3 and Proposition 1).

2.2 Beliefs

In this section, we specify the beliefs of players about the actions of others (first order beliefs), and beliefs about such beliefs (second and higher order beliefs). The set of all such beliefs is known as a *belief hierarchy*. We shall only specify *point beliefs*. Our model can be generalized to the more realistic case of a belief distribution and belief distributions over belief distributions.¹⁴ However, this does not add any new insights to our analysis but adds significant algebraic complexity. We build the required belief hierarchy in an iterative manner. We employ the convention of using the ‘player number’ in the subscript and the ‘order of the belief’ in the

¹³Our analysis goes through with a more general utility function of the trustor $u(Y - i + r)$; $u' > 0, u'' < 0$. But this does not add any significant insights to the analysis.

¹⁴For applications of the use of such belief distributions, using the induced beliefs design, see Khalmetski et al. (2015), Dhami et al. (2019), and Dhami et al. (2022). For a rich survey of applications using other methods of belief elicitation, see Battigalli and Dufwenberg, 2022.

superscript, so b_j^n is the n^{th} order belief, $n = 1, 2, \dots$, of player $j = 1, 2$. We shall only need beliefs upto order 2, so $n = 2$. Players have private information about their beliefs.

1. *First order beliefs*: The first order belief of the trustor (Player 1), denoted by $b_1^1(i) : [0, Y] \rightarrow [0, 3i]$, specifies, in the mind of Player 1, for each possible value of investment i , the corresponding expected return, r , from the trustee.¹⁵
2. *Second order beliefs*: The trustor's first order beliefs, $b_1^1(i)$, are not observed by the trustee. Hence, the trustee needs to form subjective beliefs about the trustor's first order beliefs $b_1^1(i)$ in order to, say, compute sequential reciprocity. The second order beliefs of the trustee (player 2), denoted by $b_2^2(i) : [0, Y] \rightarrow [0, 3i]$, specify, for any observed value of investment, i , the trustee's beliefs about the return, r , expected by the trustor.

One of the novelties of our paper is that we extend the existing theoretical literature by formally incorporating (i) identity, and (ii) priming, within belief hierarchies. Players may expect their ingroups ($s = 1$) to take different actions as compared to their outgroups ($s = 0$); and their beliefs might also be influenced by priming ($p = 1$) or the absence of priming ($p = 0$). To take account of this dependence, we denote the relevant beliefs more generally by $b_1^1(i; s, p)$ (trustor's first order beliefs) and $b_2^2(i; s, p)$ (trustee's second order beliefs).

We make the purely technical assumption that beliefs $b_1^1(i; s, p)$ and $b_2^2(i; s, p)$ are twice continuously differentiable with respect to i ; this facilitates the analysis. We now make three plausible assumptions on beliefs and we successfully test all three in our empirical analysis.

Assumption 1. (Responsiveness of beliefs to investment): *The response of the trustor's first order beliefs, $b_1^1(i; s, p)$, and the trustee's second order beliefs, $b_2^2(i; s, p)$, to a change in investment, i , is non-negative.*

$$(i) \frac{\partial b_1^1(i; s, p)}{\partial i} \geq 0, (ii) \frac{\partial b_2^2(i; s, p)}{\partial i} \geq 0; i \in [0, Y], s \in \{0, 1\}, p \in \{0, 1\}. \quad (2.3)$$

Assumption 1(i) requires that when the trustor sends a higher investment, i , the trustor expects a higher return from the trustee. Assumption 1(ii) requires that the trustee believes that the trustor does not expect a lower return when the trustor sends a higher investment. Underlying this assumption appears to lie a 'shared understanding' among the players that others are reciprocal. This is supported by our data.

Assumption 2. (First order beliefs of trustor, identity, and priming): *We make the following assumptions on the first order beliefs of the trustor, $b_1^1(i; s, p)$:*

$$\begin{cases} (i) b_1^1(i; 0, p) \leq b_1^1(i; 1, p), i \in [0, Y], p \in \{0, 1\}. \\ (ii) b_1^1(i; 1, 0) \leq b_1^1(i; 1, 1), i \in [0, Y]. \end{cases} \quad (2.4)$$

From the first row of (2.4), for any level of investment and priming, the trustor expects an ingroup trustee ($s = 1$) to return a higher amount than an outgroup trustee ($s = 0$). From the second row of (2.4), primed trustors ($p = 1$) expect their ingroup trustee ($s = 1$) to return even more relative to unprimed trustors ($p = 0$).

¹⁵Before the trustor sends his investment, i , Player 2, the trustee, has a first order belief, b_2^1 , about the amount the trustor will send. However, this belief does not play any role in our analysis, so we omit it.

Assumption 3. (Second order beliefs of trustee, identity, and priming) *We make the following assumptions on the second order beliefs of the trustee, $b_2^2(i; s, p)$*

$$\begin{cases} (i) b_2^2(i; 0, p) \leq b_2^2(i; 1, p), i \in [0, Y] p \in \{0, 1\}, \\ (ii) b_2^2(i; 1, 0) - b_2^2(i; 0, 0) \leq b_2^2(i; 1, 1) - b_2^2(i; 0, 1). \end{cases} \quad (2.5)$$

From the first row of (2.5), for any level of investment and priming, the trustee believes that an ingroup trustor ($s = 1$) expects a higher return as compared to an outgroup trustor ($s = 0$). The second row of (2.5) gives a *marginal identity effect*: Primed trustees ($p = 1$) believe that trustors expect even greater return differences between ingroup and outgroups, relative to unprimed trustees ($p = 0$).¹⁶

2.3 Consistency of beliefs and actions

Our beliefs-based model is in the class of models of psychological game theory, because the belief hierarchies directly enter into the utility function (Geanakoplos et al. (1989), Battigalli and Dufwenberg (2009), and Battigalli and Dufwenberg (2022)). In such models, players play their best response to their beliefs and there is mutual consistency of beliefs and actions. If players have the relevant information, then best response to beliefs is not controversial. However, the bulk of the evidence shows that ‘consistency between beliefs and equilibrium actions’ required in variations of sequential Nash equilibrium does not hold in the early rounds of most games; nor is there any guarantee that it holds in games that are repeated and learning is allowed (Dhami, 2019, Vol. 4; Dhami, 2020, Vol. 5). For this reason, as in models of non-equilibrium beliefs (e.g., level-k models, cognitive hierarchy models, evidential equilibrium, and models of cursed equilibrium) we do not require the mutual consistency of beliefs and actions.¹⁷ We clarify this further in our formal definitions below.

3 Computation of sequential reciprocity

In this section, we compute the conditional reciprocity term R in the utility function of the trustee, (2.2). The convention is that the trustor is Player 1 and the trustee is Player 2. We define the sequential reciprocity term for Player 2 (trustee) as

$$R = k_{21} \hat{k}_{12}, \quad (3.1)$$

where k_{21} is the *kindness of Player 2 to Player 1, as perceived by Player 2*; and \hat{k}_{12} is the *kindness of Player 1 to Player 2, as perceived by Player 2*. If Player 1 is perceived to be kind ($\hat{k}_{12} > 0$),

¹⁶All our results go through if we assumed, in addition, that for any level of priming, $p \in \{0, 1\}$, individuals place relatively more weight on reciprocity towards ingroup members as compared to outgroup members, i.e., $\lambda(0, p) \leq \lambda(1, p)$. However, since we do not directly measure the parameter λ , we do not impose this assumption. It does not add anything to our analysis.

¹⁷For useful surveys of the evidence, see, Mauersberger and Nagel (2018), and Dhami (2019, Vol. 4). In particular, Bellemare et al. (2011) show that there is a lack of consistency between actions, first-order beliefs, and second-order beliefs. See also Section 9 in Battigalli and Dufwenberg (2022) for a critical discussion of the solution concepts in psychological games and a recognition of the importance of non-equilibrium beliefs. For applications that do not require consistency between beliefs and actions, see Khalmetski et al. (2015), Dhami et al. (2019), and Dhami et al. (2022).

then by reciprocating the kindness ($k_{21} > 0$), Player 2 increases utility. Similarly, Player 2's utility can be increased by reciprocating perceived unkindness ($\widehat{k}_{12} < 0$) with unkindness ($k_{21} < 0$). This is the sense in which reciprocity is conditional. Reciprocity is sequential because the trustee observes the choice of the trustor before computing R .

In Proposition 1 below, we compute the reciprocity term, R , in (3.1), using the Dufwenberg and Kirchsteiger (2004) definition of reciprocity in sequential games.¹⁸ This requires defining, for each player, the 'equitable payoff' of a player, which is a weighted average of the maximum payoff (with weight γ) and the minimum payoff (with weight $1 - \gamma$) that can arise to a player from the actions of the other player. Any payoff above the equitable payoff indicates a 'kind' action by the other player; conversely payoffs below the equitable payoff are perceived to be 'unkind.'

Payoffs depend on the actions taken by the players but some of the actions might be unobserved. For instance, player 1 (the trustor) does not know the trustee's return decision, r , at the time of choosing the investment, i , but has first order beliefs, $b_1^1(i; s, p)$, about the return decision. Player 2 (the trustee) does not observe $b_1^1(i; s, p)$, but has beliefs about $b_1^1(i; s, p)$; these are player 2's second order beliefs, $b_2^2(i; s, p)$ and they play an important role in determining reciprocity.

Proposition 1. *Using the Dufwenberg and Kirchsteiger (2004) definition of reciprocity in sequential games, the reciprocity term, R , in (3.1) is given by*

$$R = R(i, r, b_2^2) = (r - 3\gamma i) [(3i - r) - (3\gamma Y - b_2^2(i; s, p))], \quad (3.2)$$

where $\gamma \in [0, 1]$ is the weight accorded to the maximum possible payoff of a player in the computation of the equitable payoff of the player.

From (3.2), sequential conditional reciprocity is the product of two terms.

1. The term $k_{21} = r - \gamma 3i$: The kindness of player 2 to player 1, as perceived by player 2, is higher, the higher is the amount returned to the trustor, r , relative to the fraction γ of the maximum possible return, $3i$.
2. The term $\widehat{k}_{12} = (3i - r) - (3\gamma Y - b_2^2(i; s, p))$. This term is positive, i.e., player 1 is kind to player 2, as perceived by player 2, if

$$(3i - r) > (3\gamma Y - b_2^2(i; s, p)). \quad (3.3)$$

From (3.3), the kindness of player 1 (as perceived by player 2) is higher, the higher is the investment, i , sent by player 1 relative to the amount returned, r . The equitable payoff of the trustee depends negatively on the trustor's first order belief, b_1^1 , of the amount to be returned, r , by the trustee (see the proof of Proposition 1).¹⁹ The trustee does not

¹⁸The kindness functions in Rabin (1993) and Dufwenberg and Kirchsteiger (2004) are related in spirit, although the specifications are slightly different. For a discussion of the alternative concepts, see Dufwenberg and Kirchsteiger (2019).

¹⁹The higher is the amount returned by the trustee, the lower is the trustee's material payoff in all states of the world, including in the computation of the equitable payoff.

observe b_1^1 , but has second order beliefs, b_2^2 , about b_1^1 , which are used as a proxy for b_1^1 . A decrease in the equitable payoff increases the difference between the actual and the equitable payoff of the trustee, increasing the kindness of player 1, as perceived by player 2. Hence, R is increasing in b_2^2 . This is a testable implication, and we successfully test it.

4 Solution to the trust game

4.1 The trustee's optimization problem

Substituting (3.2) into (2.2), we can rewrite the utility function of the trustee as

$$V(i, r; s, p) = v(3i - r) + \lambda [(r - \gamma 3i)(3(i - \gamma Y) - r + b_2^2(i; s, p))]. \quad (4.1)$$

In the spirit of backward induction, we first solve the trustee's optimization problem, followed by the trustor's optimization problem.

Definition 1. (*Psychological best response of the trustee*) In a psychological best response, the trustee chooses the optimal amount to return $r \in [0, 3i]$, for each possible observed investment $i \in [0, Y]$ sent by the trustor, in order to maximize V defined in (4.1), conditional on the trustee's second order beliefs, b_2^2 .²⁰

Using Definition 1, the trustee chooses the optimal return, conditional on the trustee's beliefs, and a given level of investment, i , that has already been chosen by the trustor.

$$r^* \in \operatorname{argmax} V(i, r; s, p) = v(3i - r) + \lambda [(r - 3\gamma i)(3(i - \gamma Y) - r + b_2^2(i; s, p))]; r \in [0, 3i]. \quad (4.2)$$

In (4.2), i is already chosen by the trustor, hence, $b_2^2(i; s, p)$ is a particular value of the trustee's second order belief corresponding to i , conditional on s, p . Differentiating (4.2), we have

$$\frac{\partial V}{\partial r} = -v'(3i - r) + \lambda [3i(1 + \gamma) - 2r - 3\gamma Y + b_2^2(i; s, p)]. \quad (4.3)$$

The two terms on the RHS of (4.3) give the marginal effects of an increase in a unit of return, r , by the trustee. The first term is the marginal disutility arising from one less unit of consumption. The second term is the marginal effect on conditional reciprocity.²¹

Proposition 2. (*Existence of a solution*) (a) A unique solution to the optimization problem of the trustee in (4.2) exists and it is given by $r^*(i, s, p)$; we term this the 'reaction function' of the trustee. If $\lambda = 0$, then we have the corner solution $r^* = 0$.

²⁰In Definition 1, we have not imposed mutual consistency of beliefs and actions (see the discussion in Section 2.3. Mutual consistency of beliefs and actions would have required that the optimal return, $r = r^*$, chosen by the trustee must equal the (i) first order belief of the trustor, b_1^1 , and (ii) the second order belief of the trustee, b_2^2 , so that $r^* = b_1^1 = b_2^2$. This is rejected by the empirical evidence (see Section 2.3 for the references). By contrast, in classical game theory mutual consistency of actions and beliefs of all orders (order $n \rightarrow \infty$) is essential.

²¹In the absence of reciprocity, ($\lambda = 0$), or if the marginal effect on reciprocity is negative, the RHS of (4.3) is negative, so $r^* = 0$. However, the empirical evidence shows that $r^* > i > 0$, i.e., the second term on the RHS in (4.3) is strictly positive; for a survey, see Dhami (2019, Vol. II). Models of other-regarding preferences, e.g., the Fehr-Schmidt (1999) model also predict $r^* > 0$.

Proposition 2 shows that the presence of conditional reciprocity ($\lambda > 0$) is a necessary condition for an interior solution. In light of the empirical evidence, the interesting case is $\lambda > 0$. We now state the testable predictions of our model for the trustee.

Proposition 3. (*Comparative statics*) Suppose that $\lambda > 0$.

- (a) (*Conditional reciprocity*) At an interior solution, the trustee's optimal choice r^* is strictly increasing in i and λ . If Assumption 1(ii) holds then r^* is strictly increasing in b_2^2 .
- (b) (*Absolute ingroup/outgroup differences*) If Assumption 3(i) holds, then trustees return a relatively higher amount, r^* , to ingroup trustors for all $p \in (0, 1)$, i.e., $r^*(i, 1, p) \geq r^*(i, 0, p)$.
- (c) (*Marginal identity effects*) Suppose that Assumption 3(ii) holds. When trustees are primed, the difference in amounts returned to their ingroup and outgroup trustors is greater, relative to the case where they are not primed, i.e.,

$$r^*(i; 1, 1) - r^*(i; 0, 1) \geq r^*(i; 1, 0) - r^*(i; 0, 0).$$

Proposition 3a describes the comparative static implications of conditional reciprocity. The optimal amount returned by the trustee is increasing in trustor's investment, i , the trustee's reciprocity parameter λ , and in the second order belief of the trustee about the expectation of the return by the trustor (size of b_2^2). None of these effects would not arise in the absence of conditional reciprocity ($\lambda = 0$). This critically identifies the channels through which priming and social identity influence the return decision of the trustee, i.e., through $b_2^2(i; s, p)$. If Assumption 1(ii) holds (and we show in Section 7, it does), then this must be the channel through which this effect works. This also highlights the importance of using rigorous theoretical predictions as a basis for experiments, as opposed to constructing plausible hypotheses based on pure introspection and intuition.

Proposition 3b shows that, for a fixed level of priming, larger amounts are returned by the trustee to ingroup trustors as compared to outgroup trustors. Proposition 3c shows that the marginal effect identified in Proposition 3b is stronger in magnitude when trustees are primed, relative to when they are not primed. The predictions in Proposition 3b,c required Assumptions 3(i), (ii) to hold, and we empirically verify these assumptions in Section 7.

4.2 The trustor's optimization problem

We begin with the definition of a psychological best response of the trustor.

Definition 2. (*Psychological best response of the trustor*) In a psychological best response, the trustor chooses the optimal level of investment $i \in [0, Y]$, in order to maximize U , defined in (2.1), conditional on the trustor's first order beliefs, $b_1^1(i; s, p)$, about the amount returned by the trustee for each possible level of investment.²²

²²We do not impose mutual consistency of beliefs and actions, which requires plugging in the optimal reaction function of the trustee, r^* , from Proposition 2, into the trustor's optimization problem. However, r^* depends, in turn, on the trustee's second order beliefs, b_2^2 that are unobserved by the trustor. Hence, the trustor would need to form third order beliefs b_1^3 (beliefs about b_2^2). Mutual consistency of beliefs and actions requires that $r^* = b_1^1 = b_2^2 = b_1^3$. The evidence (see Section 2.3, and footnote to Definition 1) shows this is unreasonable.

Using Definition 2, and (2.1), the trustor's optimization problem is

$$i^*(s, p) \in \operatorname{argmax} U(i, b_1^1(i, s, p)) = Y - i + b_1^1(i, s, p). \quad (4.4)$$

Differentiating (4.4), we get

$$\frac{\partial U(i, s, p)}{\partial i} = -1 + \frac{\partial b_1^1(i, s, p)}{\partial i}. \quad (4.5)$$

Equation (4.5) shows the marginal effects of a change in investment. The first term on the RHS is the marginal disutility of giving up a unit of consumption. The second term is the marginal benefit arising from the extra return expected from the trustee; recall from Assumption 1(i), we have that $\frac{\partial b_1^1}{\partial i} > 0$, which we empirically verify in Section 7.

Proposition 4. (*Comparative statics*) Consider the trustor's optimization problem in (4.4).

- (a) (*Existence of solution*) For all $s \in \{0, 1\}$, $p \in \{0, 1\}$, and $i \in [0, Y]$, if $\frac{\partial^2 b_1^1}{\partial i^2} < 0$, then a unique solution $i^*(s, p) \in [0, 1]$ exists and it is the solution to $\frac{\partial b_1^1(i^*, s, p)}{\partial i} = 1$.
- (b) (*Ingroup/outgroup differences*) Suppose that the conditions stated in part (a) hold. (i) If Assumption 2(i) holds, then the trustor sends a relatively higher investment to ingroup trustees, i.e., $i^*(0, p) \leq i^*(1, p)$, for any level of priming $p \in \{0, 1\}$. (ii) If Assumption 2(ii) holds, then the trustor sends a higher investment to ingroup trustees when primed, relative to the case when not primed, i.e., $i^*(1, 0) \leq i^*(1, 1)$.
- (c) (*Marginal identity effects*) Suppose that the conditions stated in part (a) and Assumptions 2(ii), (iii) hold, and the sufficient condition $i^*(0, 1) \leq i^*(1, 0)$ holds.²³ Then, the differences between investments sent by primed trustors to ingroup and outgroup trustees is higher, relative to unprimed trustors, i.e., $i^*(1, 0) - i^*(0, 0) \leq i^*(1, 1) - i^*(0, 1)$.

Trustors send more investment to ingroups relative to outgroups (Proposition 4b) and the marginal ingroup/outgroup discrimination in investment sent by the trustor is relatively more pronounced when they are primed (Proposition 4c). The proposition reveals the transmission channel (Assumption 2(i), (ii)) through which the relevant predictions arise; Assumption 2(i), (ii) are empirically tested in Section 7.

5 Identity, priming, and optimal choice in the dictator game

An analysis of ingroup/outgroup differences is incomplete without distinguishing between *taste-based discrimination* and *statistical discrimination*. In order to do so, subjects play a dictator game in the presence of identity concerns. Following Fershtman and Gneezy (2001), we keep the dictator game as close as possible to the trust game except that player 2 is passive. Player 1, the *dictator*, has an endowment, Z . The dictator shares an amount $z \in [0, Z]$ with a passive player 2, the *receiver*, who is either an ingroup or an outgroup member. Player 1 can be primed or unprimed. The experimenter triples the amount sent to the receiver, so the receiver receives $3z$. The material payoffs of player 1 and player 2 are, respectively, $Z - z$ and $3z$.

²³This condition requires that ingroup trustees are sent larger amounts by unprimed trustors, relative to the amounts sent by primed trustors to outgroup trustees.

If the dictator has purely self-regarding preferences, then it is optimal to choose $z = 0$. Hence, we need other-regarding preferences to explain dictator giving.²⁴ We use a non-linear form of the Fehr and Schmidt (1999) model, otherwise one always gets a corner solution, while the data shows a heterogeneity of interior solutions.²⁵

In our discussion below, we assume, consistent with the evidence, that dictators do not offer a share, z , such that the passive receiver's payoff, $3z$, is higher than the dictator's payoff, $Z - z$.²⁶ The dictator's preferences are given by

$$W(z; s, p) = (Z - z) - \beta(s, p)f(Z - 4z); Z - z \geq 3z, 0 \leq \beta < 1. \quad (5.1)$$

In (15.1), the dictator derives utility from own material payoffs, $Z - z$ (first term) and disutility from advantageous payoff differences, $Z - z \geq 3z$, due to altruism (second term); β is the parameter of advantageous inequity.²⁷ When $Z - z = 3z$, the material payoffs of both players are identical, so $W = \frac{3}{4}Z$. Self-regarding preferences are given by the special case $\beta = 0$.

The function f , which captures disutility from payoff differences, is increasing and convex, $f' > 0, f'' > 0$. Thus, individuals derive disutility as payoff differences increase and the 'marginal disutility' from payoff differences increases at an increasing rate. Our assumptions on beliefs made in Section 2.2 do not apply because the second player is passive. Following the empirical evidence in Chen and Li (2009) and Dhami et al. (2021), the parameter β is influenced by social identity, hence we write it as $\beta(s, p)$, where $s, p \in \{0, 1\}$; we make two assumptions below in (15.3), (15.4).

$$(i) \beta(0, p) \leq \beta(1, p); p \in \{0, 1\}. \quad (5.2)$$

Thus, people are more altruistic towards ingroups.

$$(i) \beta(1, 0) \leq \beta(1, 1), (ii) \beta(0, 1) \leq \beta(0, 0). \quad (5.3)$$

In (15.4), priming heightens the effects of social identity in (15.3). Primed subjects are more altruistic towards ingroups (see (i)) and less altruistic towards outgroups (see (ii)).

We do not directly test for the conditions in (15.3) and (15.4). However, the preferences in (15.1), in conjunction with the conditions in (15.3) and (15.4), lead to specific predictions (see Proposition 6) that we test with data. A rejection of those predictions is also a potential rejection of (15.3) and (15.4). But we are able to confirm these predictions with the data.

The objective function of the dictator is

$$z^* \in \operatorname{argmax} W(z; s, p), z \in [0, Z], \quad (5.4)$$

²⁴One of the most important lessons from behavioral economics is that preferences are context-dependent and frame-dependent (Dhami, 2016). For instance, a stock market trader or a financial analyst might be completely self-interested in their day job, yet exhibit altruism by giving money to a homeless person on the way back home, which is a close analogue to the dictator game.

²⁵Non-linear forms of Fehr-Schmidt preferences are common in applied research (Dhami, 2019, Vol. II). We could also have introduced Fehr-Schmidt preferences for the trust game, but that would have added an extra layer of complexity to the model without matching gains in insights.

²⁶In other words, we assume that the dictator is in the domain of *advantageous inequity*, $Z - z \geq 3z$. See the supplementary section for a richer model where the possibility of $Z - z < 3z$ is allowed. However, this does not change our results, because the optimal solution is in the domain of advantageous inequity.

²⁷The restriction $\beta < 1$ incorporates evidence that people do not burn their money to reduce payoff differences (Dhami, 2019, Vol. II).

where W is defined in (15.1).

Proposition 5. *Consider the optimization problem of the dictator in (15.5)*

- (a) *There is a unique solution to the dictator's problem, $z^*(s, p) \leq \frac{Z}{4}$, for $s, p \in \{0, 1\}$.*
- (b) *Suppose that $\beta(0, p) \leq \beta(1, p)$, as in (15.3), then, $z^*(0, p) \leq z^*(1, p)$, $p \in \{0, 1\}$.*
- (c) *Suppose that $\beta(1, 0) \leq \beta(1, 1)$, and $\beta(0, 1) \leq \beta(0, 0)$, as in (15.4), then, (i) $z^*(1, 0) \leq z^*(1, 1)$, (ii) $z^*(0, 1) \leq z^*(0, 0)$.*

From Proposition 6b, dictators share a larger amount with ingroups rather than outgroups for any level of priming. From Proposition 6c, priming increases altruism towards ingroups but decreases altruism towards outgroups. If the data supports the predictions in Proposition 6b,c, then it is also consistent with the assumptions in (15.3), (15.4). The next corollary gives us another testable implication of Proposition 6c in terms of marginal identity effects.

Corollary 1. *Proposition 6c implies $z^*(1, 1) - z^*(0, 1) \geq z^*(1, 0) - z^*(0, 0)$.*

6 Experimental design

The experiments were conducted in India with 542 villagers from Bihar and Uttar Pradesh. No subject participated in the experiment more than once. The average time taken to complete the experiment was 28 minutes, and the subjects earned, on average, 260 Rupees (roughly 4.7 US dollars) including a participation fee. All subjects were paid in private after the experiment through an automated process which excluded the experimenter. The study is pre-registered; see <https://doi.org/10.1257/rct.8424>. All material payoffs in the experiment are expressed in tokens that are converted into Indian Rupees at an exchange rate of 1 token = 2 Rupees. Subjects received Rs. 100 as a show-up fee.

There are two games in the experiments: the trust game and the dictator game. The two games are in a counterbalanced order, and each subject was randomly assigned to one of the two orders. The subjects got paid the earnings in both games (plus the participation fee). In the trust game, subjects were randomly assigned to be trustors or trustees.

Each *trustor* was endowed with 60 tokens. Trustors first guessed the ingroup and outgroup trustee's possible return, r , to them if their investment, i , was at 5 possible levels (1/4, 1/3, 1/2, 2/3, and 3/4 of the endowment of 60, or respectively 15, 20, 30, 40, and 45 tokens). All elicited beliefs in the experiment were incentivized and the correct guess were paid an additional Rs. 50. These guesses, which measured the trustor's first order beliefs, were *not* informed to the trustees. Next, the trustors decided the actual amount of tokens (one of 5 possible discrete levels) to be sent to an ingroup and an outgroup trustee. Each trustor and each trustee was either a Hindu or a Muslim.

Each trustee received three times the investment sent by the trustor. Before they received this amount, trustees guessed the trustor's expectations of the trustee's return for each of the five levels of potential investment sent by the trustor (15, 20, 30, 40, and 45 tokens). These guesses, which measure the trustee's second order beliefs, were *not* informed to the trustors. Next, using the strategy method, the trustees decided the amount of tokens to return to a Hindu

and a Muslim trustor for each of the five possible levels of investment the trustor might send to them.

All subjects were in the role of the dictator in the dictator game. The subjects were endowed with 20 tokens, and they decided the amount to send to a Hindu and a Muslim recipient. They knew that the amount they sent would be tripled and then given to the passive receiver.²⁸

There are two treatments - the primed treatment and the neutral treatment. Both treatments contain a reading task, followed by a writing task where the subjects are required to write three sentences related to the topic in the reading task.²⁹

In the primed treatment subjects read the following factually correct, and readily publicly available, information before being asked to write three sentences related to the topic. “Between 2005 and 2017, a total of 9722 incidents of communal violence have occurred across the country. 1471 people have died in these incidents, and 28591 have been injured. 447 incidents of promotion of enmity on grounds of religion were reported in 2016, a 38% increase over 2014, when 323 were reported. Overall, 1,148 incidents of promoting such enmity were reported between 2014 and 2016.” Subjects were then asked to write 3 sentences on the Ayodhya verdict of 2019.³⁰ We did not offer our interpretation of the events or draw any inferences from them for the subjects.

This was followed by two priming manipulation checking questions. The first question is to select the issue they were just asked to read or write about, from a few given options. The second question is a word stem completion task.³¹ Our analysis is based on the 354 subjects with the correct answer to the first priming manipulation check question, which we take as a minimum test of understanding and attention for human subjects.

6.1 List of independent variables

We use the following variables in our econometric analysis.

Religiosity: In the post-experimental survey, and following Rohrbaugh and Jessor (1975), subjects answered eight questions related to a multidimensional religiosity measure that includes four aspects of religiosity— ritual, consequential, ideological, and experiential. There

²⁸The dictator’s transfers were given to relatively low-income participants of the relevant religion who did not participate in our study.

²⁹In the neutral treatment, subjects read the following paragraph before being asked to write three sentences related to the topic. “COVID-19 has resulted in schools shut down all across the world. Globally, over 1.2 billion children are out of the classroom. As a result, education has changed dramatically, with the distinctive rise of e-learning, whereby teaching is undertaken remotely and on digital platforms. Research suggests that online learning has been shown to increase retention of information, and take less time, meaning the changes coronavirus have caused might be here to stay.”

³⁰The Ayodhya verdict relates to a Muslim shrine whose ownership was disputed between Hindus and Muslims. This was a vexed legal case due to the difficulty of establishing historical precedence. In 2019, the Supreme Court of India settled the dispute regarding ownership of the site in favour of the Hindu litigants and gave permission for the construction of a Hindu holy temple on the disputed site. This was also a major political and electoral item on the agenda of the ruling BJP party.

³¹Subjects read 14 word stems and select the first word that immediately comes to mind. They can only select one word from two options, or write a word with the relevant stem. In the 14 word stem completion tasks, six tasks consisted of one religion-related word and one neutral word. We give two examples of religion-related words. (1) For the word stem “DIVI_” the religion-related word was “divine” and the neutral word was “divide”. (2) For the word stem “SCRIP_” the religion-related word was “scripture” and the neutral word was “scripted”. Since our experiments were run in Hindi, the word stem completion tasks were also about similar Hindi word stems. The other eight tasks consisted of two neutral words. For instance the word stem “GLA_” could either be “glass” or “glad”.

were two questions on each aspect and we used a 5-point Likert scale for the answers. The aggregated scores from this multidimensional religiosity measure range between 0 and 32 (higher scores imply greater religiosity), and are called the *religiosity scores*. We term this variable as *Religiosity*.

Investment: Investment i sent by the trustor to the trustee.

Prime: Dummy variable that equals 1 for subjects in the primed treatment, and 0 otherwise. 68% (= 240/354) subjects are randomly assigned to the primed treatment.³²

Hindu: Dummy variable that equals 1 for Hindu subjects, and 0 otherwise. 64% (= 228/354) subjects are Hindu.³³

Business: Dummy variable that equals 1 for business/economics subjects, and 0 otherwise.

Married: Dummy variable that equals 1 for married subjects, and 0 otherwise.

Experience: Dummy variable that equals 1 if the subjects have attended similar experiments before, and 0 otherwise.

Female: Dummy variable that equals 1 for female subjects, and 0 otherwise. 49% (= 173/354) subjects are female.

Age: Subject's age.

Education: Dummy variable that equals 1 if the subjects have obtained bachelor or higher degree, and 0 otherwise.

FOB: First order beliefs of the trustor (see Section 2.2).

SOB: Second order beliefs of the trustee (see Section 2.2).

We shall, when convenient, also use the variable R to denote the religion of a subject; $R = H$ for Hindu subjects and $R = M$ for Muslim subjects.

For the trustor we shall often refer to the five possible investment levels as 1, 2, 3, 4, and 5. These correspond to an investment of 15, 20, 30, 40, and 45, tokens, respectively.

7 Testing assumptions on ‘beliefs’ in the trust game

We now test our three assumptions on beliefs in Section 2.2 about (i) the first order beliefs of the trustor, b_1^1 (i.e., the trustor's expectations of the amount returned by the trustee), and (ii) the second order beliefs of the trustee, b_2^2 (i.e., the trustee's beliefs about the amount that the trustor expects back).

7.1 Testing Assumption 1

7.1.1 Testing Assumption 1(i)

Assumption 1(i) requires that when the trustor sends a higher investment, i , the trustor expects a higher return from the trustee, i.e., $\frac{\partial b_1^1(i;s,p)}{\partial i} \geq 0$ for all $s, p \in \{0, 1\}$ and for all i . We

³²In our original random sample of 542 subjects, 48.2% are in the neutral treatment and 51.8% in the primed treatment. However, we only accepted data for subjects that correctly answered tests of understanding of the experiment. This lead to unequal proportion of subjects in each treatment.

³³In the first stage, we randomly sampled 400 subjects, and 80% of the sample was Hindu (300 Hindus and 100 Muslims). This is close to the actual India-wide figure of 79.8% for Hindus in the latest available figures. In order to increase the data on Muslim subjects, in the second stage of our data collection, we targeted another 100 Muslim subjects by randomly sampling from the underlying population.

compute the Spearman correlation coefficient between b_1^1 and i , for each trustor for all $s, p \in \{0, 1\}$. The upper and the lower half of Table 1 shows, respectively, results for ingroup trustees and outgroup trustees. The symbol “+” denotes significantly positive correlation, “-” denotes significantly negative correlation, and “0” denotes insignificant correlation (all at the 5% level). The categories “+” and “0” (non-negative correlation between b_1^1 and i) are consistent with Assumption 1(i).³⁴ Except for the 4% primed-Hindu trustors facing ingroup trustees, and the 3% primed-Hindu trustors, facing outgroup trustees, that lie in the category “-”, the data are consistent with Assumption 1(i).

Table 1: Using the Spearman correlation coefficient to measure the responsiveness of trustor’s first order beliefs to different investment levels, disaggregated by social identity, and the level of priming. Number of subjects in each category are shown in brackets.

ingroup trustees	Positive (+)		Zero (0)		Negative (-)	
trustor’s religion	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	66% [61/93]	77% [34/44]	30% [28/93]	23% [10/44]	4% [4/93]	0% [0/44]
neutral	69% [35/51]	58% [14/24]	31% [16/51]	42% [10/24]	0% [0/51]	0% [0/24]
outgroup trustees	Positive (+)		Zero (0)		Negative (-)	
trustor’s religion	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	60% [56/93]	73% [32/44]	37% [34/93]	27% [12/44]	3% [3/93]	0% [0/44]
neutral	80% [41/51]	75% [18/24]	20% [10/51]	25% [6/24]	0% [0/51]	0% [0/24]

7.1.2 Testing Assumption 1(ii)

Assumption 1(ii) requires that the trustee believes that the trustor expects a higher return when the trustor sends a higher investment, i.e., $\frac{\partial b_2^2(i;s,p)}{\partial i} \geq 0$ for all levels of $s, p \in \{0, 1\}$. We proceed as in our test of Assumption 1(i) and use identical notation, except that we now use second order beliefs of the trustee, b_2^2 . The results, using the sign of the Spearman correlation in different cases, are shown in Table 2.³⁵ The categories “+” and “0” (non-negative correlation between b_2^2 and i) are consistent with Assumption 1(ii). These two categories contain 100% of the data in the neutral treatment and between 90% and 98% of the data for primed Hindu and Muslim trustees. This is consistent with Assumption 1(ii).

³⁴The numbers in Table 1 show the percentage of trustors who satisfy a given condition. For instance, 66% of Hindu trustors expect ingroup trustees to return a higher amount if they send a higher investment.

³⁵The numbers in Table 2 show the percentage of trustees who satisfy a given condition. For instance, 80% of Hindu trustees believe that ingroup trustors will expect a higher amount to be returned if they send a higher investment.

Table 2: Using the Spearman correlation coefficient to measure the responsiveness of the trustee’s second order beliefs to investment, disaggregated by social identity and the level of priming. Absolute numbers in each category are shown in brackets.

ingroup trustors	Positive (+)		Zero (0)		Negative (-)	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	80% [51/64]	72% [28/39]	13% [8/64]	26% [10/39]	7% [5/64]	2% [1/39]
neutral	75% [15/20]	95% [18/19]	25% [5/20]	5% [1/19]	0% [0/20]	0% [0/19]
outgroup trustors	Positive (+)		Zero (0)		Negative (-)	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	56% [36/64]	77% [30/39]	34% [22/64]	23% [9/39]	10% [6/64]	0% [0/39]
neutral	70% [14/20]	89% [17/19]	30% [6/20]	11% [2/19]	0% [0/20]	0% [0/19]

7.2 Testing Assumption 2

7.2.1 Testing Assumption 2(i)

Assumption 2(i) requires trustors to expect an ingroup trustee to return a higher amount compared to an outgroup trustee, i.e., $b_1^1(i, 0, p) \leq b_1^1(i, 1, p)$ for all i and for $p = \{0, 1\}$. The categories, “ingroup>outgroup”, “ingroup=outgroup”, and “ingroup<outgroup”, in Table 3, show the percentage of trustors who believe that ingroup trustees will return, respectively, higher, same, and lower amounts than outgroup trustees for ‘all’ the five possible investment levels (15, 20, 30, 40, and 45). The last category “inconclusive” shows the percentage of cases where, a subject violates the given condition, even for a single level of investment. Hence, we have fairly stringent test of the assumption. The number of subjects in each category is shown in brackets. The two rows in Table 3 differentiate between subjects in the primed treatment (first row) and the neutral treatment (second row). The cases “ingroup>outgroup” and “ingroup=outgroup” are both consistent with Assumption 2(i), which holds for 74%–92% of the subjects.

Table 3: Trustor’s first order beliefs of amounts returned by the trustee, averaged across 5 different levels of investment, and categorized by ingroup/outgroup differences.

	ingroup>outgroup		ingroup=outgroup		ingroup<outgroup		inconclusive	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	57% [53/93]	55% [24/44]	17% [16/93]	30% [13/44]	12% [11/93]	2% [1/44]	14% [13/93]	14% [6/44]
neutral	57% [29/51]	25% [6/24]	25% [13/51]	67% [16/24]	6% [3/51]	8% [2/24]	12% [6/51]	0% [0/24]

There are stark differences in the effects of priming on the first order beliefs of Hindu trustors relative to Muslim trustors. This differential effect of priming carries over to the second order

beliefs of the trustees (see Section 7.3) as well as individual actions of the trustors and trustees (Sections 8 and 11), indicating that this is a robust phenomenon in our study. This is a new finding in the literature.

1. ingroup>outgroup: In Table 3, the percentage of Hindu trustors who believe that the ingroup trustees will return more than the outgroup trustees in both the primed and the neutral treatments is identical— 57%. By contrast, 25% of Muslim trustors believe that the ingroup trustees will return more than the outgroup trustees in the neutral treatment and this percentage increases to 55% in the primed treatment; a statistically significant increase (z test, p -value < 0.01).
2. ingroup=outgroup: Two-thirds (67%) of the Muslim trustors in the neutral treatment expected identical returns from ingroup and outgroup trustees. In the primed treatment, this proportion reduces to 30%; a statistically significant reduction (z test, p -value < 0.01). However, for this category, the respective beliefs of Hindu subjects in the primed and the neutral treatments about the return expected from the trustees are 17% and 25%; these proportions are not statistically different (z test, p -value > 0.1)³⁶.

Thus, unlike Hindu trustors, once primed, Muslim trustors assign lower trustworthiness to outgroup trustees (Hindus) relative to ingroup trustees (Muslims). We have discussed potential reasons in the introduction.

The average first order belief of Muslim trustors about the return expected from Hindu trustees and Muslim trustees is respectively 29.74 and 33.15. The average first order belief of Hindu trustors about the return expected from Hindu trustees and Muslim trustees is respectively 35.1 and 30.64. For both Muslim and Hindu trustors, the average return expected from the ingroup trustees is significantly higher than that for the outgroup trustees (one-sided t test, p -value < 0.01). This conclusion also holds at each of the five investment levels, 1–5, of the trustor’s possible investment for both Hindu and Muslim trustors. The ingroup/outgroup differences are statistically significant at each investment level for Hindu trustors (one-sided t test, p -value < 0.05); for Muslim trustors the differences are insignificant for the investment level $\frac{3}{4}Y$, but only mildly significant at the remaining investment levels. These differences are likely to have arisen from Muslim trustees making little ingroup/outgroup differences when not primed.

The average first order beliefs of Hindu and Muslim trustors towards their respective ingroup trustees are not significantly different; neither are these differences significant for beliefs towards outgroup trustees (two-sided t test, p -value > 0.1). However, if we separate the trustor’s first order belief data into the primed and neutral treatments, then the results are different. The average first order belief about the expected return is higher from the ingroup trustees for both Hindu and Muslim trustors in the primed treatment, but the ingroup/outgroup differences are insignificant for the Muslim trustors in the neutral treatment (one-sided t test, p -value > 0.1). These results hold even if we separate our subjects by gender.

³⁶The pairwise comparisons of the other proportions between the primed and neutral treatments in Table 3 are insignificant at the 1% level (z test).

We conclude that Assumption 2(i) holds in our sample.

7.2.2 Testing Assumption 2(ii)

Assumption 2(ii) requires that, *ceteris-paribus*, primed trustors expect a higher return from ingroups, relative to the neutral treatment, i.e., $b_1^1(i, 1, 0) \leq b_1^1(i, 1, 1)$. This assumption is imposed at the level of each individual subject, hence, it should properly be tested in a within-subjects design. In our experiments, the subjects in the role of trustors were either in the primed treatment, or in the neutral treatment, but not both. Therefore, we cannot use a within-subjects design to test $b_1^1(i, 1, 0) \leq b_1^1(i, 1, 1)$. For this reason, we report the results of a between-subjects test. We pool the first order beliefs in the primed and the neutral treatments.

For Hindu trustors, who comprise 68% of our sample, the average first order belief in the primed treatment, $b_1^1(i, 1, 1)$, is significantly higher than that in the neutral treatment, $b_1^1(i, 1, 0)$ (one-sided t test, p -value < 0.1), as required by Assumption 2(ii). Separating the data into the five possible investment levels, 1–5, the results are still consistent with Assumption 2(ii) but are statistically insignificant (two-sided t test, p -value > 0.1). For the Muslim trustors, we find the contrary result but it is only significant at the middle levels of investment, 2, 3, 4.

7.3 Testing Assumption 3

Assumption 3 imposes restrictions on the second order beliefs of the trustee, b_2^2 , i.e., the trustee’s beliefs about the trustor’s expectations of the amount to be returned.

7.3.1 Testing Assumption 3(i)

Assumption 3(i), $b_2^2(i, 0, p) \leq b_2^2(i, 1, p)$, requires that, for any level of investment, trustees believe ingroup trustors expect a higher return than outgroup trustors, for all levels of priming, $p = 0, 1$. Table 4 shows the trustee’s second order beliefs of the trustor’s expected return, taking account of the five possible investment levels, 1–5, from the trustor. We organize the data into the same four categories that we have used earlier in Table 3.

The three categories, “ingroup>outgroup”, “ingroup=outgroup”, and “ingroup<outgroup”, show the percentage of trustees who believe that ingroup trustors expect, respectively, higher, same, and lower return than outgroup trustors at ‘all’ five possible investment levels, 1–5. The last category “inconclusive” shows the percentage of cases where the stated condition fails for even 1 out of the 5 investment levels.

Table 4: Trustee’s second order beliefs of the trustor’s expected return, categorized by ingroup/outgroup differences. Number of subjects in brackets.

	ingroup>outgroup		ingroup=outgroup		ingroup<outgroup		inconclusive	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
Primed	64%	56%	17%	26%	8%	3%	11%	15%
	[41/64]	[22/39]	[11/64]	[10/39]	[5/64]	[1/39]	[7/64]	[6/39]
Neutral	90%	5%	0%	84%	0%	0%	10%	11%
	[18/20]	[1/19]	[0/20]	[16/19]	[0/20]	[0/19]	[2/20]	[2/19]

The categories “ingroup>outgroup” and “ingroup=outgroup” are consistent with Assumption 3(i), and 81% – 90% of the subjects satisfy it; the data for 10% – 15% of the subjects is inconclusive. Consider the stark differences in the effects of priming on the second order beliefs of Hindu and Muslim trustees.

1. Ingroup>outgroup: 90% of Hindu trustees in the neutral treatment are in the category ingroup>outgroup and this number ‘falls’ to 64% in the primed treatment; a statistically significant decrease (z test, p -value< 0.01). Only 5% of Muslim trustees fall in the category ingroup>outgroup and this ‘increases’ to 56% in the primed treatment; a statistically significant increase (z test, p -value< 0.01).
2. ingroup=outgroup: In the neutral treatment, 84% of the Muslim trustees are in the category ingroup=outgroup, but this ‘falls’ to 26% in the primed treatment; a statistically significant fall (z test, p -value< 0.01). However, the percentage of Hindus in the category ingroup=outgroup is 0% in the neutral treatment, and this increases to 17% in the primed treatment; a statistically significant increase (z test, p -value< 0.01).

Thus, priming increases expectations of religious polarization in trust among Muslim trustees (as measured by their second order beliefs about the trustor’s beliefs). However, priming reduces expectations about such polarization among Hindu trustees.

The average second order belief of Muslim trustees about the return, r , expected by Muslim and Hindu trustors is, respectively, 33.07 and 30.2. For Hindu trustees, the average second order beliefs about Hindu and Muslim trustors are respectively 33.21 and 27.5. For trustees of both religions, the differences are statistically significant (one-sided t test, p -value< 0.05). This conclusion holds for each of the five investment levels, 1–5, but the differences are only highly significant for the Hindu trustees (one-sided t test, p -value< 0.05).

The Hindu and Muslim trustees’ average second order belief towards their respective ‘ingroup trustors’ are not significantly different, and the same is true for the respective ‘outgroup trustors’ (two-sided t test, p -value> 0.1).³⁷ If we separate the data for the primed and the neutral treatments, then the average second order belief of trustees is significantly higher towards ingroup trustors for both Hindu and Muslim trustees; this is also true if data are separated by gender (one-sided t test, p -value< 0.01). However, the difference between ingroups/outgroups is insignificant for Muslim trustees in the neutral treatment (two-sided t test, p -value> 0.1).

To sum up, the assumption $b_2^2(i, 0, p) < b_2^2(i, 1, p)$ is consistent with the data for Hindu and Muslim trustees in our sample.

7.3.2 Testing Assumption 3(ii)

Assumption 3(ii) requires $b_2^2(i, 1, 0) - b_2^2(i, 0, 0) \leq b_2^2(i, 1, 1) - b_2^2(i, 0, 1)$. In words: Primed trustees, relative to unprimed trustees, believe that, for a given level of investment, i , trustors expect relatively greater return difference between ingroup and outgroup trustees.

Figure 1 shows the average second order beliefs of Hindu and Muslim trustees, pooled across all 5 investment levels of the trustor. We use the 2×2 categorization between ingroup/outgroup

³⁷Recall that this is also true for the first order beliefs of trustors of both religions.

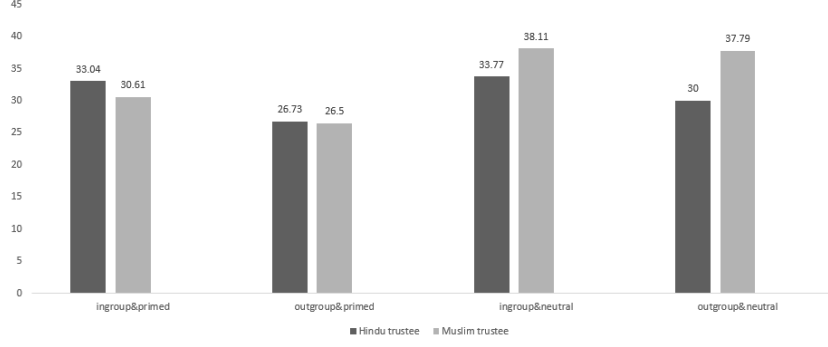


Figure 1: Average second order beliefs of Hindu and Muslim trustees towards ingroup and outgroup trustors, in the primed and neutral treatments.

trustors on the one hand and primed/neutral treatments on the other. For instance, the category ‘ingroup & primed’ denotes the beliefs of primed trustees towards ingroup trustors.

For Muslim trustees, the one-sided t test shows the average difference on the left hand side of the inequality ($b_2^2(i, 1, 0) - b_2^2(i, 0, 0) = 38.11 - 37.79 = 0.32$) is significantly lower than the right hand side of the inequality ($b_2^2(i, 1, 1) - b_2^2(i, 0, 1) = 30.61 - 26.5 = 4.11$) with p -value < 0.01 . For Hindu trustees, the one-sided t test shows the left hand side of the inequality ($33.77 - 30 = 3.77$) is significantly lower than the right hand side inequality ($33.04 - 26.73 = 6.31$) with p -value < 0.01 .³⁸

In sum, the data are consistent with Assumption 3(ii) for both Muslim and Hindu trustees.

8 Testing predictions on ‘actions’ in the trust game

8.1 Trustworthiness, social identity, and priming

In this section, we consider the predictions for the amounts returned by trustees (a measure of trustworthiness) in Proposition 3. We first give some descriptive statistics. Table 5 shows, for each of the 5 possible levels of investment sent by the trustor, the rounded average percentage of the endowment of the trustees ($3i$) that is returned back to the trustor, i.e., $r/3i$. We disaggregate the data by the religious identity of the trustee, by treatments, and by ingroups/outgroups.

The meta study by Johnson and Mislin (2011), that does not make an ingroup/outgroup distinction, shows that trustees return, on average, 37.2% of their endowment to trustors. Taking the mean level of investment that can be sent by the trustor, 30, for Hindu subjects, this percentage ranges from 33 (for primed trustees returning to outgroup trustors) to 45 (unprimed trustees returning to ingroup trustors). The corresponding percentages for Muslim trustees, for an investment level of 30, are 32 and 46.

³⁸For Muslim trustees, we also find that Assumption 3(ii) holds (one-sided t test, p -value < 0.01) separately at each of the five potential investment levels chosen by the trustor, with the only insignificant case being at the investment level of $\frac{1}{4}Y$. For Hindu trustees, for each of the five investment levels considered separately, we also confirm that Assumption 3(ii) holds (one-sided t test, p -value < 0.1); but the two insignificant cases are at the investment levels of 15 and 40, chosen by the trustor.

Table 5: Average levels of the trustee’s return, r , as a percentage of their endowment, $3i$, for each of 5 possible levels of investment (15, 20, 30, 40, 45) sent by the trustor.

Trustee	Hindu					Muslim				
	15	20	30	40	45	15	20	30	40	45
Trustor’s Investment										
Primed & Ingroup	48	46	41	35	35	44	40	37	32	33
Primed & Outgroup	41	36	33	29	29	38	36	32	29	28
Neutral & Ingroup	56	51	45	43	38	47	46	46	44	49
Neutral & Outgroup	46	44	40	36	35	49	50	48	45	50

8.1.1 Social identity and trustworthiness (Proposition 3b)

Proposition 3b predicts that trustees will return a relatively higher amount, r^* , to ingroup trustors, i.e., $r^*(i, 1, p) \geq r^*(i, 0, p)$, for any level of priming, p . In Table 6, we present a categorization that is similar to Tables 3 and 4, and requires that the condition in any given category holds for all five investment levels, 1–5, from the trustor.³⁹ The number of subjects satisfying each category is in the brackets.

Table 6: Trustees’ choices of amounts returned to trustors, categorized by ingroup/outgroup differences. Number of subjects in brackets.

	ingroup>outgroup		ingroup=outgroup		ingroup<outgroup		inconclusive	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	66% [42/64]	59% [23/39]	16% [10/64]	18% [7/39]	6% [4/64]	8% [3/39]	12% [8/64]	15% [6/39]
neutral	65% [13/20]	5% [1/19]	25% [5/20]	63% [12/19]	0% [0/20]	27% [5/19]	10% [2/20]	5% [1/19]

The categories, “ingroup>outgroup” and “ingroup=outgroup”, are consistent with Proposition 3(b). There is a high degree of conformity with our prediction, particularly in the primed treatment; results are inconclusive, on average, for between 5% – 15% of the subjects. We now describe significant differences in the effects of priming on Hindu and Muslim trustees that mirror similar differences in their second order beliefs, considered above.

1. The percentage of Hindu trustees who returned more to ingroup trustors (ingroup>outgroup) in the primed and neutral treatments is nearly identical, 66% and 65%, and insignificantly different (z test, p -value> 0.1). There is also no statistical difference in the percentage of Hindu trustees in the category ‘ingroup=outgroup.’
2. Only 5% of the Muslim trustees in the neutral treatment returned more to the ingroup trustors (ingroup>outgroup), but this increases to 59% in the primed treatment; a statistically significant increase (z test, p -value< 0.01). A significant majority, 63%, of the

³⁹The first category “ingroup>outgroup”, means that the amount returned by trustees to the ingroup trustor is higher than that to the outgroup trustor (i.e., $r^*(i, 1, p) > r^*(i, 0, p)$). The second category “ingroup=outgroup” implies $r^*(i, 1, p) = r^*(i, 0, p)$. The third category “ingroup<outgroup” implies $r^*(i, 1, p) < r^*(i, 0, p)$. The final category, labeled “inconclusive” reports subjects who fail a given condition for any 1 out of 5 investment levels.

Muslim trustees in the neutral treatment returned exactly the same to ingroup and outgroup trustors (ingroup=outgroup), but this falls to 18% in the primed treatment; a statistically significant fall (z test, p -value < 0.01).

Therefore, as with second order beliefs, priming hardly alters the behavior of Hindu trustees, but it significantly influences the behavior of Muslim trustees. The results also shows how actions are reflected in the underlying beliefs.

The average amount returned by Muslim trustees to Hindu and Muslim trustors is respectively 33.12 and 35.33, while the average amount returned by Hindu trustees to Hindu and Muslim trustors is respectively 36.08 and 30.16. For both Muslim and Hindu trustees, the average amount returned to ingroup trustors is significantly higher than the average amount returned to outgroup trustors (one-sided t test, p -value < 0.1). An identical conclusion can be drawn at each of the five investment levels, 1–5, of the trustor’s possible investment. However, the differences are highly significant only for Hindu trustees (one-sided t test, p -value < 0.01) but not the Muslim trustees.

The Hindu trustees’ and the Muslim trustees’ average return towards their respective ingroup trustors is not significantly different. Muslim trustees, return significantly higher amounts to outgroup trustors as compared to the Hindu trustees (one-sided t test, p -value < 0.05). In the primed treatment, the return of Muslim trustees is significantly higher to the ingroup trustors than outgroup trustors on average (one-sided t test, p -value < 0.01), but this difference is insignificant in the neutral treatment (two-sided t test, p -value > 0.1). The average return of Hindu trustees towards ingroup/outgroup trustors is significantly different in both treatments and favors ingroups (one-sided t test, p -value < 0.01). If we separate the trustees’ return data by gender, then the average return is higher to the ingroup trustors for both genders of Hindu and Muslim trustees (one-sided t test, p -value < 0.01), but insignificant for the Muslim female trustees (two-sided t test, p -value > 0.1).

To sum up, our theoretical prediction $r(i, 1, p) > r(i, 0, p)$ is supported by the data.

8.1.2 Priming and trustworthiness (Proposition 3c)

Proposition 3c on marginal identity effects predicts that the social identity effects (ingroup vs outgroup) become relatively more pronounced when subjects are primed, i.e., $r^*(i; 1, 1) - r^*(i; 0, 1) \geq r^*(i; 1, 0) - r^*(i; 0, 0)$. The average return of Hindu and Muslim trustees in the primed and the neutral treatments is depicted in Figure 2, separated by the two treatments, primed and neutral.

For Muslim trustees, a one-sided t test shows the average difference on the left hand side of the inequality ($r^*(i; 1, 1) - r^*(i; 0, 1) = 32.07 - 28.22 = 3.85$) is significantly higher than the difference on the right hand side ($r^*(i; 1, 0) - r^*(i; 0, 0) = 42.02 - 43.18 = -1.16$) with p -value < 0.01. For Hindu trustees, the average difference on the left hand side of the inequality, 6.14, is higher than the difference on the right hand side, 5.22; however, the two-sided t test shows that the two differences are insignificantly different (p -value = 0.190).

In sum, our data are consistent with Proposition 3c.

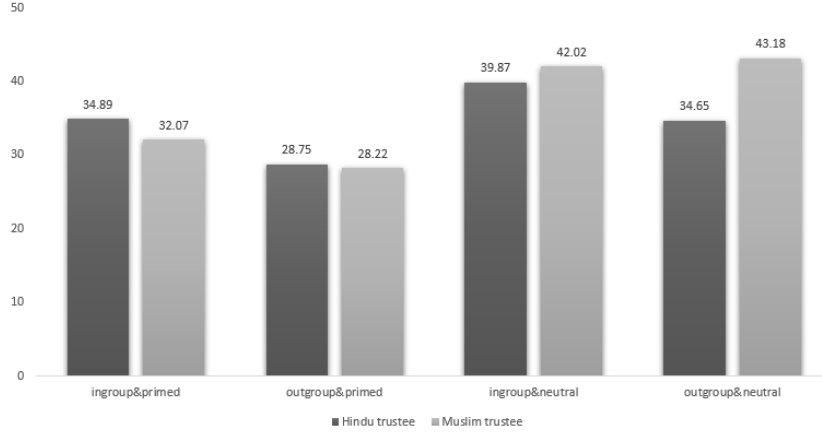


Figure 2: Average return of Hindu and Muslim trustees in the primed and neutral treatments

8.2 Trust, social identity, and priming

In this section, we provide empirical tests of Proposition 4, which predicts the comparative static effects on the optimal amount invested by trustors (a measure of trust). Table 7 shows the rounded percentage of Hindu and Muslim trustors in the various treatments who send each of the 5 categories of investment.

Table 7: Percentage of trustors who send each of the 5 different categories of investment, 15, 20, 30, 40, 45.

Trustor Investment	Hindu					Muslim				
	15	20	30	40	45	15	20	30	40	45
Primed & Ingroup	8	17	42	31	2	2	23	57	16	2
Primed & Outgroup	18	41	26	11	4	14	50	32	2	2
Neutral & Ingroup	10	12	49	23	6	4	42	33	17	4
Neutral & Outgroup	14	47	27	10	2	4	42	50	0	4

From the meta study by Johnson and Mislin (2011), we know that the mean amount sent by trustors in the meta study is 50.2% of their endowment. However, this is an average across studies that did not have social identity concerns. In our study, the mean amount is 30 out of an endowment of 60. For Hindu trustees, the percentage of subjects choosing the mean amount, 30, ranges from 26% for primed trustors who invest in outgroup trustees, to 49% for unprimed trustors who invest in ingroup trustees. The corresponding figures for Muslim trustees are 32% and 33%.

8.2.1 Trust and social identity (Proposition 4b)

Proposition 4b predicts that trustors send a relatively higher investment to ingroup trustees relative to outgroup trustees, i.e., $i^*(1, p) \geq i^*(0, p)$, for any level of priming $p \in \{0, 1\}$. The three categories are ingroup>outgroup ($i^*(1, p) > i^*(0, p)$); ingroup=outgroup ($i^*(1, p) = i^*(0, p)$); ingroup<outgroup ($i^*(1, p) < i^*(0, p)$). The number of subjects in each category is shown in brackets. The choices in the first two categories are consistent with the predictions of Proposition

4b, and this accounts for more than 85% of the subjects.

Table 8: Trustor’s choices of investment, categorized by ingroup/outgroup differences. Number of subjects in brackets.

	ingroup>outgroup		ingroup=outgroup		ingroup<outgroup	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	55%	59%	32%	32%	13%	9%
	[51/93]	[26/44]	[30/93]	[14/44]	[12/93]	[4/44]
neutral	57%	17%	29%	79%	14%	4%
	[29/51]	[4/24]	[15/51]	[19/24]	[7/51]	[1/24]

There are significant differences between Hindu and Muslim trustors, in the effects of priming on the amounts invested; this mirrors similar differences in their first order beliefs.

1. In Table 8, 57% of Hindu trustors send more investment to ingroup trustees in the natural treatment and this figure falls slightly to 55% when they are primed; the two figures are not statistically different (z test, p -value > 0.1). The percentage of Hindu trustors who send the same amounts to ingroup and output trustees (ingroup=outgroup) is also statistically indistinguishable in the primed and the neutral treatments (32% and 29%).
2. In the neutral treatment, only 17% Muslim trustors send more investment to their ingroups and this figure increases to 59% in the primed treatment; which is significantly higher (z test, p -value < 0.01). In the neutral treatment, 79% of Muslim trustors send identical amounts to ingroup and outgroup trustees, but this figure falls to 32% when primed; this is a statistically significant decrease (z test, p -value < 0.01).

Therefore, as in the case of beliefs and the return behavior of trustees, Muslim subjects are relatively more sensitive to priming. When unprimed, they exhibit little difference in trust between ingroup and outgroup trustees. But when primed, they trust their ingroup significantly more. Hindu trustors on the other hand are not significantly influenced by priming.

The average investment of Muslim trustors, sent to Hindu and Muslim trustees, is respectively 24.34 and 28.68, while the average investment of Hindu trustors, sent to Hindu and Muslim trustees, is respectively 30.59 and 24.76. For both Muslim and Hindu trustors, the average investment sent to their respective ingroups is significantly higher than the average investment sent to their respective outgroups (one-sided t test, p -value < 0.01).

The average investment of Hindu and Muslim trustors towards their respective ingroup trustees (two-sided t test, p -value > 0.1) is not significantly different (respectively, 30.59 and 28.68). Average investment towards their respective outgroups for Hindu trustors and Muslim trustors, respectively, 24.76 and 28.68, is also not significantly different (two-sided t test, p -value > 0.1). When we separate the data into the primed and the neutral treatments, the average investment sent to ingroup trustees is higher for both Hindu and Muslim trustors. However, the difference is insignificant for the Muslim trustors in the neutral treatment (two-sided t test, p -value > 0.1). For both males and females, the average investment is higher to ingroup trustees for Hindu and Muslim trustors (one-sided t test, p -value < 0.01).

In sum, our theoretical prediction $i^*(1, p) \geq i^*(0, p)$ is consistent with the data.

8.2.2 Trust and priming (Proposition 4c)

Proposition 4c, on marginal identity effects, predicts that the difference between the investments sent by the trustor to ingroup and outgroup trustees is higher for primed trustors, relative to unprimed trustors, i.e., $i^*(1, 1) - i^*(0, 1) \geq i^*(1, 0) - i^*(0, 0)$.

For Muslim trustors, the one-sided t -test shows that the average difference on the left hand side of the inequality ($i^*(1, 1) - i^*(0, 1) = 2.93 - 2.29 = 0.64$) is significantly higher than the difference on the right hand side ($i^*(1, 0) - i^*(0, 0) = 2.75 - 2.58 = 0.17$) with p -value < 0.01 . For Hindu trustors, the one-sided t test shows the average difference on the left hand side of the inequality (0.61) is insignificantly different from the right hand side (0.65) with p -value > 0.1 .

Thus, we find that the marginal identity effect on trust for Muslim trustors is significant, but it is insignificant for Hindu trustors.

9 Social identity, trust, and discrimination in the dictator game

9.1 Social identity and dictator transfers

From Proposition 6b, dictators are predicted to share a larger amount with ingroup receivers relative to outgroup receivers, for any level of priming, i.e., $z^*(1, p) \geq z^*(0, p)$, $p \in \{0, 1\}$.

Table 9 summarizes data from individual subjects. The categories are as in Table 8; e.g., the first category “ingroup>outgroup”, means the dictator’s transfer to the ingroup recipient is relatively higher. The number of subjects in each category is given in the brackets. Recall that dictators make only two decisions: transfer to a Hindu recipient, and to a Muslim recipient. The first two categories in Table 9 (“ingroup>outgroup” and “ingroup=outgroup”) are consistent with $z^*(1, p) \geq z^*(0, p)$, and satisfy Proposition 6b. This is true for a little over 90% of our subjects, on average.

Table 9: Dictator’s choices of transfers to the recipients by religious identity and categorized by ingroup/outgroup differences. Number of subjects in brackets.

	ingroup>outgroup		ingroup=outgroup		ingroup<outgroup	
	Hindu	Muslim	Hindu	Muslim	Hindu	Muslim
primed	57% [53/93]	57% [25/44]	35% [33/93]	34% [15/44]	8% [7/93]	9% [4/44]
neutral	49% [25/51]	25% [6/24]	37% [19/51]	71% [17/24]	14% [7/51]	4% [1/24]

We continue to find the same differential effects of priming on Hindu and Muslim subjects as we have found for first order/second order beliefs, trust, and trustworthiness, in the trust game.

1. In Table 9, the percentage of Hindu dictators transferring either more, or the same, to the ingroup recipients (ingroup>outgroup and ingroup=outgroup) is statistically indistinguishable in the primed and neutral treatments (z test, p -value > 0.1).

2. The vast majority (71%) of the Muslim dictators in the neutral treatment transfer exactly the same amounts to ingroup and outgroup recipients, but this percentage falls to 34% in the primed treatment; a statistically significant fall (z test, p -value < 0.01). Similarly, only 25% of the Muslim dictators in the neutral treatment transfer more to the ingroups, but this proportion increases to 57% in the primed treatment; a statistically significant increase (z test, p -value < 0.01).

The average transfer to an ingroup recipient is significantly higher relative to an outgroup recipient for Hindu dictators ($9.73 > 7.5$) and Muslim dictators ($7.88 > 6.48$); in both cases the one-sided t test, p -value < 0.01. The non-parametric sign test also strongly confirms the significantly higher median transfer to ingroup recipients (p -value < 0.01). This result is unchanged, and significant at the 5% level, if we separated data by (i) primed versus neutral treatments, or (ii) gender.

9.2 Priming and dictator transfers

We use the implication of Proposition 6c given in Corollary 2. Thus, we want to show that priming subjects increases the marginal effects of transfers towards ingroup recipients relative to outgroup recipients, i.e., $z^*(1, 1) - z^*(0, 1) \geq z^*(1, 0) - z^*(0, 0)$.

For Muslim dictators, the one-sided t test shows the average difference on the left hand side of the inequality ($z^*(1, 1) - z^*(0, 1) = 8.35 - 6.53 = 1.82$) is significantly higher than the difference on the right hand side ($z^*(1, 0) - z^*(0, 0) = 6.98 - 6.4 = 0.58$) with p -value < 0.01. For Hindu dictators, the one-sided t test shows the average difference on the left hand side of the inequality ($9.89 - 7.36 = 2.53$) is significantly higher than the right hand side ($9.37 - 7.81 = 1.56$) with the p -value < 0.05. Thus, our results conform with Proposition 6c, and priming sharpens the marginal effects of ingroup/outgroup differences.

10 Statistical discrimination and taste-based discrimination

In a trust game, statistical discrimination arises as follows (Fershtman and Gneezy, 2001). Consider two religions (or ethnic groups), R1 and R2, and the following two conditions. (1) Trustors of both religions, R1 and R2, are observed to send uniformly lower amounts of investment to trustees of one particular religion, say, trustees of religion R2. This may indicate that trustors of religion R1 and R2 have common/similar beliefs that members of religion R2 return lower amounts. Hence, it is rational, for trustors of both religions, to send R2 trustees a lower amount. (2) In a dictator game, members of both religions do not discriminate between recipients of religion R1 and R2. These two conditions furnish evidence that there is no taste based discrimination (from the dictator game), but there is statistical discrimination (from the trust game). Fershtman and Gneezy (2001) report statistical discrimination for Israeli subjects who belonged to one of two ethnic groups, Ashkinazi Jews (R1) and Eastern Jews (R2). We show in this section that, for Hindu and Muslim subjects in India, there is no statistical discrimination, and the only discrimination is taste based discrimination.

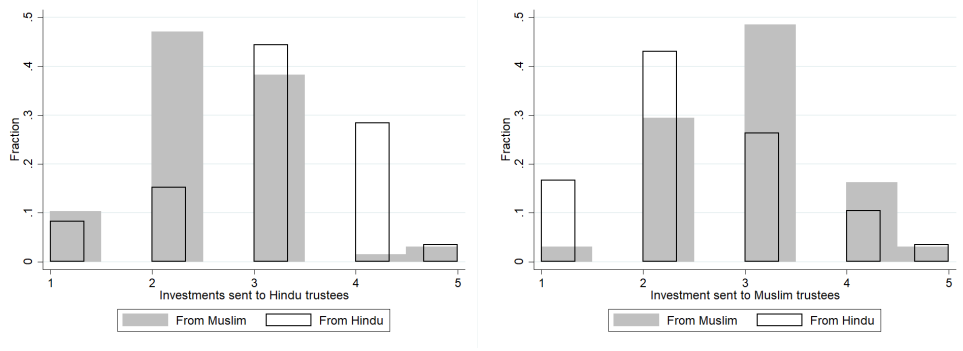


Figure 3: Investment sent to Hindu and Muslim trustees by trustors from both religions.

Figure 3 shows the pattern of trust toward Hindu and Muslim trustees, as captured by the 5 different levels of investment, 1-5, sent by the trustors in a trust game. In Section 7, 8, we document strong evidence of ingroup favouritism and outgroup discrimination in the trust game.⁴⁰ Along similar lines, it is straightforward to show that there is no statistical discrimination in terms of trustworthiness either, and there are strong ingroup/outgroup effects in the amounts returned by trustees (see also Section 8.1).⁴¹

We have already shown in Section 9 that the dictator transfers to ingroup recipients are significantly higher than to outgroup recipients for both Hindu and Muslim dictators. This conclusion remains unchanged if we separate data in terms of gender or priming. Thus, the results from the dictator game also provide strong evidence for taste-based discrimination. But we find no evidence for statistical discrimination.

11 Determinants of actions and beliefs in the trust game

We now consider the determinants of (1) trust (amount invested by the trustor), (2) trustworthiness (amount returned by trustee), (3) the first order beliefs of trustors, b_1^1 , and (4) the second order beliefs of trustees, b_2^2 . Since 64% of our subjects are Hindus, and since Hindus and Muslims differ in several respects, we present the their results separately.

11.1 Determinants of trustee's second order beliefs, b_2^2 .

Table 10 reports the determinants of the trustee's second order beliefs, b_2^2 (denoted by SOB), about the trustor's expected return. The independent variables are defined in Section 6. We run OLS regressions with robust standard errors. The Tobit model produces results similar, thus, for ease of interpretation, we present the OLS estimates. Models 1, 3 are the basic models, and Models 2, 4, respectively, add the interaction term Prime \times Ingroup, which is positive in both models. In all models in Table 10, trustees believe that trustors expect a higher return if

⁴⁰The average transfer from Hindu and Muslim trustors to Hindu trustees was respectively 30.59 and 24.34, and the former is significantly greater than the latter (one-sided t test, p -value < 0.01). The average transfer from Hindu and Muslim trustors to Muslim trustees was respectively 24.76 and 28.68, and the former is significantly lower than the latter (one-sided t test, p -value < 0.01). Thus members of one religion do not receive uniformly lower investments.

⁴¹For supporting statistical analysis, see the supplementary section

Table 10: Determinants of trustee's second order beliefs, b_2^2 .

OLS	Model 1	Model 2	Model 3	Model 4
Trustee's religion	Hindu		Muslim	
Investment	0.78*** [0.048]	0.78*** [0.048]	0.93*** [0.049]	0.93*** [0.049]
Prime	-1.30 [0.982]	-2.57* [1.441]	-9.54*** [1.197]	-11.44*** [1.652]
Ingroup	5.71*** [1.021]	3.77*** [1.396]	2.87*** [1.033]	0.32 [2.402]
Prime \times Ingroup		2.54 [1.885]		3.79 [2.599]
Religiosity	0.05 [0.207]	0.05 [0.207]	0.65*** [0.220]	0.65*** [0.219]
Age	0.11 [0.082]	0.11 [0.082]	-0.20*** [0.068]	-0.20*** [0.068]
Female	1.31 [1.068]	1.31 [1.067]	1.29 [1.102]	1.29 [1.100]
Married	-7.76*** [1.540]	-7.76*** [1.540]	-2.72 [2.228]	-2.72 [2.227]
Education	1.31 [1.626]	1.31 [1.625]	10.80*** [2.055]	10.80*** [2.063]
Constant	5.90* [3.181]	6.87* [3.200]	9.78*** [3.345]	11.06*** [3.518]
Adjusted R^2	0.303	0.304	0.505	0.507
No. of obs.	840	840	580	580

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

trustors choose higher investment (see coefficient on Investment); this anticipation of reciprocity is relatively greater among Muslim trustees. This is consistent with Assumption 1(ii). Trustees deemed more religious (higher scores on the multidimensional religiosity measure) have higher second order beliefs than those with lower religiosity, i.e., they believe that trustors expect greater trustworthiness. However, the coefficients are only significant for Muslim trustees, and are close to zero for Hindu trustees. Married Hindu trustees have significantly lower SOB relative to unmarried Hindu trustors. However, marital status is insignificant for Muslim trustors. More educated Muslim trustees also believe that trustors expect significantly higher return; but, education has insignificant effects on the SOB of Hindu trustees. Age is insignificant for Hindu trustees, while older Muslim trustees have significantly lower SOB.

The following calculations refer to the richer Models 2 and 4 for trustees of each religion. Denote by $E[SOB | R, p, s]$ the expected SOB, b_2^2 , of a trustee, conditioned on religion ($R = H, M$), priming ($p = 0, 1$), and social identity ($s = 0, 1$). We calculate below, the average

differences between ingroups and outgroups.

$$E[SOB | R = H, p = 1, s = 1] - E[SOB | R = H, p = 1, s = 0] = 6.31. \quad (11.1)$$

$$E[SOB | R = M, p = 1, s = 1] - E[SOB | R = M, p = 1, s = 0] = 4.11. \quad (11.2)$$

$$E[SOB | R = H, p = 0, s = 1] - E[SOB | R = H, p = 0, s = 0] = 3.77. \quad (11.3)$$

$$E[SOB | R = M, p = 0, s = 1] - E[SOB | R = M, p = 0, s = 0] = 0.32. \quad (11.4)$$

From (11.1), primed Hindu trustees believe that Hindu trustors expect 6.31 units more, on average, relative to Muslim trustors. From (11.2), the average ingroup/outgroup differences in SOBs for primed Muslim trustees are positive, but smaller than Hindu trustees. From (11.3), unprimed Hindu trustees believe that Hindu trustors expect 3.77 units more, on average, relative to Muslim trustors. From (11.4), unprimed Muslim trustees believe that Muslim trustors expect 0.32 units more, on average, relative to Hindu trustors. Thus, they expect a lower degree of religious polarization in the behavior of trustors, relative to Hindu trustees.

The findings in (11.1)-(11.4) satisfy Assumption 3(i). Subtracting (11.3) from (11.1) we get $6.31 - 3.77 = 2.54 > 0$, and subtracting (11.4) from (11.2) we get $4.11 - 0.32 = 3.79 > 0$; in conjunction this satisfies Assumption 3(ii). The marginal identity effects in beliefs are slightly stronger for Muslim trustees.

11.2 Determinants of trustees' optimal return

Table 11 reports the determinants of the trustee's actual choice of return, r (dependent variable), using OLS regressions with robust standard errors (results for the Tobit model are similar). To the basic models (Models 1 and 3), we added the interaction terms Prime \times Ingroup, Prime \times SOB, Prime \times SOB \times Ingroup, and Prime \times Religiosity to obtain richer models (Models 2 and 4). The three-way interaction is not significant for trustees of any religion, but Prime \times Ingroup, and Prime \times SOB are significant for Muslim trustees, while Prime \times Religiosity is significant for Hindu trustees.

From Table 11, trustees return more to the trustors if (i) trustors choose higher investment, and (ii) if their second order belief, SOB (b_2^2), is higher, i.e., they believe the trustor expects a higher return. The effects of SOB on the amount returned are similar for Hindu and Muslim trustees, but the reciprocity effects of higher investment are stronger for Muslim trustees. These findings conform with the prediction of Proposition 3a that arises on account of sequential reciprocity.⁴² Trustees who received higher scores on the multidimensional religiosity measure, return more than those with lower religiosity; the effects are relatively stronger for Hindu trustees. Older, female, unmarried, less educated, trustees also return higher amounts to the trustors. However, the effects of marital status, age, and education are much higher in magnitude and highly significant for Hindu trustees only (age is significant only at the 10% level for Muslim trustees).

⁴²The positive effect of second order beliefs on the amount return may also arise due to *guilt-aversion*, as in Dharmi et al. (2019). However, our experiments are not designed to separate the effects of reciprocity and guilt.

Table 11: Determinants of trustee's return.

OLS	Model 1	Model 2	Model 3	Model 4
Trustee's religion	Hindu		Muslim	
Investment	0.30*** [0.047]	0.29*** [0.047]	0.39*** [0.066]	0.47*** [0.068]
Prime	-5.33*** [1.122]	12.86** [5.944]	-6.22*** [1.006]	-3.07 [5.551]
Ingroup	2.37** [0.919]	2.80 [1.794]	0.42 [0.903]	-1.38 [1.574]
SOB	0.62*** [0.045]	0.64*** [0.067]	0.62*** [0.051]	0.72*** [0.049]
Religiosity	0.48*** [0.171]	1.99*** [0.497]	0.61*** [0.168]	0.38 [0.392]
Age	0.37*** [0.087]	0.39*** [0.087]	0.08 [0.051]	0.08* [0.048]
Female	0.70 [1.094]	0.85 [1.075]	4.30 [0.940]	3.95*** [0.900]
Married	-4.57*** [1.498]	-4.75*** [1.468]	-2.42 [1.599]	-1.85 [1.505]
Education	-2.68*** [1.018]	-2.15** [1.018]	-2.76 [2.029]	0.83 [2.169]
Prime \times Ingroup		-0.61 [3.006]		5.35* [2.979]
Prime \times SOB		-0.01 [0.092]		-0.32*** [0.083]
Prime \times SOB \times Ingroup		-0.00 [0.088]		-0.06 [0.088]
Prime \times Religiosity		-1.79*** [0.520]		0.43 [0.412]
Constant	-4.63 [3.195]	-20.21*** [5.431]	-2.97 [3.081]	-5.97 [5.776]
Adjusted R^2	0.550	0.556	0.670	0.691
No. of obs.	840	840	580	580

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Next, we examine the effects of social identity and priming. Denote by $E[r | R, p, s]$ the expected amount returned, r , by a trustee, conditioned on religion ($R = H, M$), priming ($p = 0, 1$), and social identity ($s = 0, 1$). We calculate below, the average differences between ingroups and outgroups based on Models 2, 4.

$$E[r | R = H, p = 1, s = 1] - E[r | R = H, p = 1, s = 0] = 2.19. \quad (11.5)$$

$$E[r | R = M, p = 1, s = 1] - E[r | R = M, p = 1, s = 0] = 3.91. \quad (11.6)$$

$$E[r | R = H, p = 0, s = 1] - E[r | R = H, p = 0, s = 0] = 2.80 \quad (11.7)$$

$$E[r | R = M, p = 0, s = 1] - E[r | R = M, p = 0, s = 0] = -1.38. \quad (11.8)$$

From (11.5), primed Hindu trustees return 2.19 units more, on average, to Hindu trustors relative to Muslim trustors. From (11.6), the ingroup/outgroup differences in return for primed Muslim trustees are higher than those for Hindu trustees. From (11.7), unprimed Hindu trustees return 2.80 units more, on average, to ingroup trustors. From (11.8) unprimed Muslim trustees return 1.38 units less, on average, to ingroup trustors; thus, when unprimed, they act in a more trustworthy manner towards Hindu trustors.

The findings in (11.5)-(11.8) satisfy Proposition 3b. Subtracting (11.7) from (11.5) we get $2.19 - 2.80 = -0.61 < 0$, and subtracting (11.8) from (11.6) we get $3.91 + 1.38 = 5.29 > 0$; the second of these results is consistent with the prediction of Proposition 3c and the first is not. Priming does not sharpen the return differences between ingroups and outgroups for Hindu trustees, but it does so significantly for Muslim trustees; this is consistent with our direct tests of the assumptions above.

11.3 Determinants of the trustor's first order beliefs

Table 12 reports the determinants of the trustor's first order beliefs, b_1^1 (FOB) of the amount returned by the trustee, using OLS regressions with robust standard errors (results are similar with the Tobit model).

Models 1, 3 are the basic models while Models 2, 4 add the extra interaction term Prime \times Ingroup. Trustors believe trustees will return more if trustors choose higher investment. They appear to anticipate, correctly, the conditional reciprocity of the trustees; Muslim trustors expect slightly higher reciprocity. More religious trustors (higher scores on the multidimensional religiosity measure) have relatively higher first order beliefs, i.e., they expect greater trustworthiness. However, this effect is only significant for Muslim trustors. Hindu female trustors expect lower return relative to male Hindu trustors; while no gender effects in FOB are found for Muslim trustors. Married Hindu trustors expect lower return relative to unmarried Hindu trustors; but there is no effect of marital status on the FOBs of Muslim trustors. More educated Hindu trustors expect significantly higher return, but education does not play a significant role in the FOB of Muslim trustors.

We now consider the findings on the effect of identity and priming on the trustor's FOB based on Models 2, 4. Denote by $E[FOB | R, p, s]$ the FOB, b_1^1 , of a trustor, conditioned on

Table 12: Determinants of the trustor's first order beliefs about the amount returned by the trustee, b_1^1 .

OLS	Model 1	Model 2	Model 3	Model 4
Trustor's religion	Hindu		Muslim	
Investment	0.85*** [0.039]	0.85*** [0.039]	0.97*** [0.050]	0.97*** [0.050]
Prime	1.99** [0.825]	1.33 [1.176]	-4.69*** [1.239]	-5.52*** [1.652]
Ingroup	4.47*** [0.842]	3.62*** [1.010]	3.41*** [1.039]	2.33 [1.817]
Prime \times Ingroup		1.32 [1.554]		1.67 [2.214]
Religiosity	0.25 [0.164]	0.25 [0.164]	1.48*** [0.244]	1.48*** [0.244]
Age	0.09** [0.040]	0.09** [0.040]	-0.15** [0.072]	-0.15** [0.072]
Female	-2.43*** [0.916]	-2.43*** [0.916]	-0.71 [1.214]	-0.71 [1.214]
Married	-2.87* [1.482]	-2.87* [1.482]	1.55 [1.277]	1.55 [1.280]
Education	4.80*** [1.107]	4.80*** [1.108]	1.44 [2.272]	1.44 [2.269]
Constant	1.31 [2.334]	1.74 [2.333]	-8.66** [3.473]	-8.12** [3.541]
Adjusted R^2	0.289	0.288	0.445	0.445
No. of obs.	1440	1440	680	680

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

religion ($R = H, M$), priming ($p = 0, 1$), and social identity ($s = 0, 1$).

$$E[FOB \mid R = H, p = 1, s = 1] - E[FOB \mid R = H, p = 1, s = 0] = 4.94. \quad (11.9)$$

$$E[FOB \mid R = M, p = 1, s = 1] - E[FOB \mid R = M, p = 1, s = 0] = 4.00. \quad (11.10)$$

$$E[FOB \mid R = H, p = 0, s = 1] - E[FOB \mid R = H, p = 0, s = 0] = 3.62. \quad (11.11)$$

$$E[FOB \mid R = M, p = 0, s = 1] - E[FOB \mid R = M, p = 0, s = 0] = 2.33. \quad (11.12)$$

From (11.9), primed Hindu trustors believe that Hindu trustees will return 4.94 units more, on average, relative to Muslim trustees. From (11.10), the ingroup/outgroup differences in beliefs for Muslim trustors are positive but slightly smaller than those of Hindu trustors. From (11.11), unprimed Hindu trustors believe that Hindu trustees will return 3.62 units more, on average, relative to Muslim trustees. From (11.12), unprimed Muslim trustors believe that Muslim trustees will return 2.33 units more, on average, relative to Hindu trustees.

The findings in (11.9)-(11.12) jointly satisfy Assumptions 2(i), (ii).

11.4 Determinants of the trustor’s optimal investment

To find the determinants of the trustor’s choice of investment, we report the results of the ordered probit model in Table 13. The dependent variable is the trustor’s choice of investment from among 5 different investment levels. Models 1, 3 are the basic models and Models 2, 4 are the richer models that include several interaction terms. We denote the cut points of the ordered probit model by μ_1 , μ_2 , μ_3 , and μ_4 . Recall that trustors choose a separate investment level for ingroup/outgroup trustees, and we also elicit the trustors’ first order beliefs for each of the five potential investment levels for both ingroup and outgroup trustees.

From Table 13, trustors with higher FOB are likely to send higher investments. It is more likely that trustors send higher investments to the religious ingroup trustees, as predicted in Proposition 4b. The variable Prime is not significant for Muslim trustors, however, the variable, Prime \times Ingroup, is significant for Hindu and Muslim trustors. The interaction term Prime \times FOB \times Ingroup is quite small in magnitude, but more significant for Muslim trustors. The more educated Hindu trustors are likely to send lower investment to the trustees; but education is insignificant in explaining the behavior of Muslim trustors. Religiosity is not significant in Models 2, 4. Gender, marital status, and age, are not significant factors in explaining the behavior of trustors.

In Table 14, we report selected marginal effects from Table 13; see supplementary files for more details on the table. For Hindu subjects, we use Model 2 and for Muslim subjects, we use Model 4. An increase in the FOB makes it more likely that trustors send higher investment levels; and a decrease in the probability that they send lower investment levels. Religiosity, by itself, has insignificant marginal effects. But when primed, more religious Hindu trustors are significantly less likely to send higher amounts; these effects are insignificant for Muslim trustors.

The magnitudes of the two interactions, Prime \times FOB and Prime \times FOB \times Ingroup, are insignificant. The marginal effects of Prime \times Ingroup are significant statistically and have relatively high magnitudes. For instance, primed Hindu trustors are 22% more likely to send high investment levels 3, 4, 5 to their ingroups relative to unprimed trustors. However, this figure is even higher, 37%, for Muslim trustors. In the absence of any priming, Hindu trustors are 10% more likely to send high investment levels 3, 4, 5 to their ingroups, relative to outgroups. The corresponding figure for Muslim trustors is insignificant. These results tie in with our earlier discussion in Section 7 where we showed that (i) priming has relatively greater effects on Muslim trustors, as compared to Hindu trustors, and (ii) that in the absence of priming, Muslim trustors show relatively muted ingroup/outgroup differences relative to Hindus.

For Hindu trustors, there are negative and significant marginal effects of higher education on the probability of sending higher investment levels (investment levels 3, 4, 5); correspondingly, the marginal effects on lower investments levels (1 and 2) are significantly positive. For Muslim trustors, the the marginal effects of education are insignificant. For Hindu trustors, all other demographic variables are insignificant. For Muslim trustors, all demographic variables are insignificant.

Table 13: Determinants of trustor's investment.

Ordered probit	Model 1	Model 2	Model 3	Model 4
Trustor's religion	Hindu		Muslim	
FOB	0.03*** [0.006]	0.06*** [0.011]	0.08*** [0.011]	0.08*** [0.016]
Prime	-0.03 [0.158]	1.78** [0.826]	0.29 [0.313]	0.46 [0.860]
Ingroup	0.51*** [0.102]	0.32** [0.153]	0.40*** [0.152]	0.03 [0.154]
Religiosity	-0.04 [0.037]	0.07 [0.056]	-0.11*** [0.040]	-0.05 [0.064]
Age	0.01 [0.009]	0.01 [0.009]	-0.01 [0.018]	-0.01 [0.019]
Female	0.13 [0.163]	0.14 [0.164]	0.30 [0.295]	0.27 [0.301]
Married	-0.26 [0.287]	-0.29 [0.290]	-0.59* [0.348]	-0.55 [0.355]
Education	-0.55* [0.213]	-0.53** [0.218]	-0.21 [0.387]	-0.15 [0.377]
Prime \times Ingroup		0.71** [0.339]		1.41*** [0.510]
Prime \times FOB		-0.02* [0.013]		0.02 [0.022]
Prime \times FOB \times Ingroup		-0.02* [0.008]		-0.03** [0.014]
Prime \times Religiosity		-0.12* [0.068]		-0.08 [0.085]
μ_1	-0.50 [0.417]	1.12 [0.712]	-1.27 [0.649]	-0.85 [0.922]
μ_2	0.64 [0.398]	2.29 [0.710]	0.58 [0.615]	1.10 [0.866]
μ_3	1.81 [0.406]	3.49 [0.723]	2.76 [0.690]	3.30 [0.871]
μ_4	3.08 [0.438]	4.79 [0.741]	4.20 [0.959]	4.64 [1.087]
No. of obs.	288	288	136	136

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 14: Marginal effects of the Model 2 and Model 4 in Table 13.

Investment level	FOB		Ingroup		Prime		Prime \times Ingroup		Prime \times Religiosity	
	Model 2	Model 4	Model 2	Model 4	Model 2	Model 4	Model 2	Model 4	Model 2	Model 4
1	-0.01*** [0.002]	-0.01*** [0.003]	-0.05** [0.025]	-0.00 [0.014]	-0.29** [0.136]	-0.04 [0.082]	-0.12** [0.055]	-0.13*** [0.050]	0.02* [0.011]	0.01 [0.008]
2	-0.01*** [0.002]	-0.01*** [0.003]	-0.05** [0.025]	-0.00 [0.027]	-0.27** [0.130]	-0.08 [0.149]	-0.11** [0.053]	-0.25*** [0.090]	0.02* [0.011]	0.01 [0.015]
3	0.00*** [0.001]	0.01*** [0.003]	0.02* [0.014]	0.00 [0.025]	0.14** [0.066]	0.08 [0.142]	0.05** [0.026]	0.23*** [0.073]	-0.01* [0.005]	-0.01 [0.014]
4	0.01*** [0.002]	0.01*** [0.001]	0.06** [0.028]	0.00 [0.011]	0.33** [0.151]	0.03 [0.063]	0.13** [0.064]	0.10*** [0.038]	-0.02* [0.012]	-0.01 [0.007]
5	0.00*** [0.001]	0.00* [0.001]	0.02* [0.010]	0.00 [0.004]	0.11* [0.061]	0.01 [0.026]	0.04* [0.022]	0.04 [0.027]	-0.01* [0.005]	-0.00 [0.003]

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

12 Determinants of dictator's transfer

Table 15 reports the determinants of the dictators' choice of transfers using OLS regressions with robust standard errors (the Tobit model produces similar results).

Table 15: Determinants of dictator's choices of transfer to the receiver in the dictator game.

OLS	Model 1	Model 2	Model 3	Model 4
Dictator's religion	Hindu		Muslim	
Prime	-0.10 [0.320]	-0.58 [0.418]	0.57 [0.516]	-0.05 [0.673]
Ingroup	2.22*** [0.304]	1.56*** [0.521]	1.40*** [0.461]	0.58 [0.868]
Prime \times Ingroup		0.96 [0.641]		1.25 [1.020]
Religiosity	0.02 [0.065]	0.02 [0.064]	-0.03 [0.089]	-0.03 [0.088]
Age	0.01 [0.022]	0.01 [0.022]	0.01 [0.026]	0.01 [0.026]
Female	0.09 [0.331]	0.09 [0.330]	0.34 [0.471]	0.34 [0.471]
Married	0.93* [0.492]	0.93* [0.491]	0.41 [0.625]	0.41 [0.623]
Education	0.01 [0.364]	0.01 [0.362]	1.52 [0.939]	1.52 [0.943]
Constant	6.19*** [0.867]	6.52*** [0.880]	5.66*** [1.309]	6.07*** [1.362]
Adjusted R^2	0.107	0.110	0.033	0.035
No. of obs.	454	454	252	252

Notes: The robust standard errors are in the brackets. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The only significant variable for both Hindu and Muslim trustors is the dummy variable, Ingroup. Dictators of both religions make significantly higher transfers to ingroup receivers. In addition, married Hindu dictators transfer more to the receivers; but marital status is not significant for Muslim dictators. We examine next the effects of priming and social identity based on Model 2 for Hindu dictators and Model 4 for Muslim dictators. Denote by $E[t | R, p, s]$ the expected transfers t , by the dictator, conditioned on religion ($R = H, M$), priming ($p = 0, 1$), and social identity ($s = 0, 1$).

$$E[t | R = H, p = 1, s = 1] - E[t | R = H, p = 1, s = 0] = 2.52. \quad (12.1)$$

$$E[t | R = M, p = 1, s = 1] - E[t | R = M, p = 1, s = 0] = 1.83. \quad (12.2)$$

$$E[t | R = H, p = 0, s = 1] - E[t | R = H, p = 0, s = 0] = 1.56 \quad (12.3)$$

$$E[t | R = M, p = 0, s = 1] - E[t | R = M, p = 0, s = 0] = 0.58. \quad (12.4)$$

From (12.1), primed Hindu dictators gave 2.52 units more, on average, to Hindu receivers relative

to Muslim receivers. From (12.2), primed Muslim dictators give 1.83 units more to their ingroup receivers relative to the outgroup receivers. From (12.3), unprimed Hindu dictators give 1.56 units more, on average, to Hindu receivers relative to Muslim receivers. From (12.4), unprimed Muslim dictators also give 0.58 units more, on average, to Muslim receivers relative to Hindu receivers.

The findings in (12.1)-(12.4) show that the predictions of Proposition 6b and Corollary 2 hold for dictators of both religions. Subtracting (12.3) from (12.1), we get $2.52 - 1.56 = 0.96 > 0$; while subtracting (12.4) from (12.2) we get $1.83 - 0.58 = 1.25 > 0$. This is as predicted in Corollary 2.

13 Conclusions

In this paper, we offer a comprehensive analysis of trust, trustworthiness, reciprocity, and prosociality for Hindu and Muslim subjects in India.

We contribute to the theoretical literature on identity economics in several ways. We show how the belief hierarchies of people (first and second order beliefs) influence the actions they take, which in turn determine trust, trustworthiness, reciprocity, and prosociality. We make explicit assumptions on the roles of identity and priming on belief hierarchies that fundamentally guide the design of our experiments. These assumptions are directly confirmed when they are stringently tested against the evidence. Furthermore, the predictions of the model, that are based on these assumptions, are confirmed by the evidence. This provides strong justification for the transmission mechanisms that we have identified in our model. Our methods are novel, foundational, and portable.

We find that priming subjects has asymmetric effects on Hindu and Muslim subjects. Priming makes little difference to the beliefs and actions of Hindu subjects in the trust and the dictator game. By contrast, priming significantly sharpens ingroup/outgroup polarization among the beliefs and actions of Muslim subjects. But when unprimed, Muslim subjects do not engage in as much ingroup/outgroup polarization as Hindu subjects. The average levels of trust, trustworthiness, and prosociality of Hindu and Muslim subjects towards their respective ingroups and outgroups are comparable.

There are important differences between Hindu and Muslim subjects in the significance of the roles played by religiosity, education, marital status, and age. As predicted by our theory, the *marginal identity effects* (differences between ingroups vs outgroups in beliefs and actions) of primed subjects are stronger relative to unprimed subjects. We demonstrate that for both Hindu and Muslim subjects, there is no statistical discrimination in the trust game (either at the level of trust or trustworthiness) and all observed ingroup/outgroup differences are taste based. Our multidimensional religiosity measure shows that more religious people are more trusting and more trustworthy and this is reflected in the underlying beliefs.

Acknowledgements

We are grateful to the Centre for Social and Behaviour Change (CBSE), Ashoka University, the National Natural Science Foundation of China (72003100), and Fellowship of China Postdoctoral Science Foundation (2020M670616) for the funding of this research. The experiments were run by the CSBC Behavioural Lab in the Field. We are grateful to Yan Chen for suggestions on the experimental design. Declarations of Interests: None.

14 Appendix: Proofs

Proof of Proposition 1: The material payoff of player 1 is $Y - i + \gamma 3i$. Denote the *equitable material payoff* for player 1 by $E_1(i, r)$.

$$E_1(i, r) = \gamma \max \{(Y - i + r), r \in [0, 3i]\} + (1 - \gamma) \min \{(Y - i + r), r \in [0, 3i]\}; \gamma \in [0, 1]. \quad (14.1)$$

Thus, $E_1(i, r)$ is a weighted average of the maximum and the minimum material payoffs that player 2 can guarantee player 1 through the return decision, $r \in [0, 3i]$. The maximum material payoff to Player 1 arises when $r = 3i$, and the minimum when $r = 0$. Thus, we can rewrite (14.1) as $E_1 = Y - i + \gamma 3i$. Define k_{21} (kindness of player 2 to player 1, as perceived by player 1) as the difference between the material and equitable payoffs of player 1.

$$k_{21}(r) = (Y - i + r) - (Y - i + \gamma 3i) = r - \gamma 3i. \quad (14.2)$$

Thus, player 2 is kind to player 1 if through the choice of the return decision, r , player 1 receives a material payoff greater than the equitable payoff. Otherwise player 2 is unkind to player 1.

Analogous to (14.1), and noting that the material payoff of player 2 is $3i - r$, the equitable (material) payoff of player 2, E_2 , is

$$E_2 = \gamma \max \{(3i - r), i \in [0, Y]\} + (1 - \gamma) \min \{(3i - r), i \in [0, Y]\}; \gamma \in [0, 1]. \quad (14.3)$$

The maximum material payoff to player 2 arises when the trustor chooses $i = Y$, and the minimum when the trustor chooses $i = 0$. Thus, we can rewrite (14.3) as $E_2 = 3\gamma Y - r$. However, when player 1 chooses i , the return decision, r , by player 2 is unobserved. Hence, player 1 must use their first order beliefs $b_1^1(i; s, p)$ in estimating the likely return, r . Thus, the equitable payoff of player 2, from the point of view of player 1 is

$$E_2 = 3\gamma Y - b_1^1(i; s, p). \quad (14.4)$$

The kindness of player 1 to player 2, as perceived by player 2, \hat{k}_{12} , is given by the difference between the actual and equitable payoffs of player 2, thus

$$\hat{k}_{12} = (3i - r) - (3\gamma Y - b_1^1(i; s, p)). \quad (14.5)$$

However, player 2 does not observe the first order beliefs of player 1, b_1^1 . Hence, player 2 must form second order beliefs about the first order beliefs of player 1, conditional on receiving an

investment, i . This belief, for the investment level, i , is given by $b_2^2(i, s, p)$. Hence, we can rewrite (14.5) as

$$\widehat{k}_{12} = (3i - r) - (3\gamma Y - b_2^2(i; s, p)). \quad (14.6)$$

Substituting (14.2) and (14.6) in (3.1) we get

$$R = R(i, r, b_2^2) = (r - 3\gamma i) ((3i - r) - (3\gamma Y - b_2^2(i; s, p))). \blacksquare \quad (14.7)$$

Proof of Proposition 2: V in (4.2) is a continuous function of r defined on the closed and bounded set $[0, 3i]$. Hence, it reaches a maximum at some $r \in [0, 3i]$. Differentiating (4.3) with respect to i , $\frac{\partial^2 V}{\partial r^2} = v'' - 2\lambda < 0$. Thus, V is a strictly concave function of r , hence, the maximum value is unique. Evaluating (4.3) at $\lambda = 0$ we get $\frac{\partial V}{\partial r} = -v'(3i - r) < 0$, hence, $r^* = 0$. \blacksquare

Proof of Proposition 3: We first state an intermediate result. An interior solution, $r^* \in (0, 1)$, can be found by setting the RHS of (4.3) equal to zero. Let $x \in \{i, \lambda, b_2^2\}$. Then, using the implicit function theorem gives

$$\frac{\partial r^*}{\partial x} = \left(-\frac{\partial^2 V}{\partial r^2} \right)^{-1} \frac{\partial^2 V}{\partial r \partial x}; \quad x \in \{i, \lambda, b_2^2\}. \quad (14.8)$$

Using (14.8),

$$\text{sign} \frac{\partial r^*}{\partial x} = \text{sign} \frac{\partial^2 V}{\partial r \partial x}; \quad x \in \{i, \lambda, b_2^2\}.$$

(a) For an interior solution, $r^* \in (0, 1)$, and using (4.3), we must have

$$\frac{\partial V}{\partial r} = -v'(3i - r) + \lambda \left[[3(i - \gamma Y) - r + b_2^2(i; s, p)] - (r - \gamma 3i) \right] = 0. \quad (14.9)$$

The first term on the RHS is strictly negative, so the second must be strictly positive. From Proposition 2, at an interior solution, $\lambda > 0$, hence,

$$(3(i - \gamma Y) - r + b_2^2) - (r - \gamma 3i) > 0. \quad (14.10)$$

Differentiating (4.3) with respect to i we get

$$\frac{\partial^2 V}{\partial i \partial r} = -3v''(3i - r) + \lambda \left[3(1 + \gamma) + \frac{\partial b_2^2(i; s, p)}{\partial i} \right] > 0. \quad (14.11)$$

Since $r \leq 3i$, and $v'' < 0$, the first term on the RHS of (14.11) is strictly positive. Assumption 1(ii) guarantees that the second term is positive, hence, $\frac{\partial^2 V}{\partial r \partial i} > 0$. Thus, using the implicit function theorem

$$\text{sign} \frac{\partial r^*}{\partial i} = \text{sign} \frac{\partial^2 V}{\partial r \partial i} > 0.$$

Differentiating (4.3) with respect to λ , we get

$$\frac{\partial^2 V}{\partial \lambda \partial r} = [3(i - \gamma Y) - r + b_2^2(i; s, p)] - (r - \gamma 3i) > 0. \quad (14.12)$$

The sign of (14.12) follows using (14.10). Using the implicit function theorem,

$$\text{sign} \frac{\partial r^*}{\partial \lambda} = \text{sign} \frac{\partial^2 V}{\partial \lambda \partial r} > 0.$$

Differentiating (4.3) with respect to b_2^2 , we get $\frac{\partial^2 V}{\partial b_2^2 \partial r} = \lambda > 0$. Using the implicit function theorem, we get

$$\text{sign} \frac{\partial r^*}{\partial b_2^2} = \text{sign} \frac{\partial^2 V}{\partial b_2^2 \partial r} > 0.$$

(b) We now explore the effects of social identity, s , holding fixed the level of priming, p . Since $s \in \{0, 1\}$ is a binary variable we cannot directly differentiate with respect to s . We have for small Δs

$$\frac{\Delta r^*}{\Delta s} \approx \frac{dr^*}{db_2^2} [b_2^2(i; 1, p) - b_2^2(i; 0, p)] \geq 0. \quad (14.13)$$

From (a) we know that $\frac{dr^*}{db_2^2} > 0$. From Assumptions 3(i), the term in the square brackets on the RHS of (14.13) is non-negative, hence, $\frac{\Delta r^*}{\Delta s} \geq 0$.

(c) We are now interested in the change in the marginal effect shown in (14.13) between primed and unprimed subjects. Using (14.13), we have for small Δs

$$\left(\frac{\Delta r^*}{\Delta s} \right)_{p=1} - \left(\frac{\Delta r^*}{\Delta s} \right)_{p=0} \approx \frac{dr^*}{db_2^2} [(b_2^2(i; 1, 1) - b_2^2(i; 0, 1)) - (b_2^2(i; 1, 0) - b_2^2(i; 0, 0))] \geq 0. \quad (14.14)$$

From (a) we know that $\frac{dr^*}{db_2^2} > 0$. From Assumption 3(ii) the term in the square brackets on the RHS of (14.13) is non-negative, hence, the RHS of (14.14) is non-negative. Thus, $r^*(i; 1, 1) - r^*(i; 0, 1) \geq r^*(i; 1, 0) - r^*(i; 0, 0)$. ■

We need the next Lemma for the proof of Proposition 4.

Lemma 1. *Assumption 2(i) implies $\frac{\partial b_1^1(i^*(0, p), 0, p)}{\partial i} \leq \frac{\partial b_1^1(i^*(1, p), 1, p)}{\partial i}$.*

Proof of Lemma 1: Suppose that Assumption 2(i) holds, so that $b_1^1(i; 0, p) < b_1^1(i; 1, p)$. Since the assumption holds for all levels of investment, it follows that when $i \in (0, Y)$, $b_1^1(i + \Delta i; 0, p) < b_1^1(i + \Delta i; 1, p)$, where Δi is a small change in i .⁴³ Subtracting the first inequality from the second, dividing by Δi , and taking limits on both sides as $\Delta i \rightarrow 0$, we get $\frac{\partial b_1^1(i, 0, p)}{\partial i} \leq \frac{\partial b_1^1(i, 1, p)}{\partial i}$. ■

Proof of Proposition 4: (a) Suppose that $\frac{\partial^2 b_1^1}{\partial i^2} < 0$ for all i . Differentiating (4.5) again with respect to i , we get $\frac{\partial^2 U}{\partial i^2} = \frac{\partial^2 b_1^1}{\partial i^2} < 0$. Thus, U is a strictly concave function of i on the closed and bounded interval $[0, Y]$. It follows that there is a unique maximizer, i^* , that can be found by setting the RHS of (4.5) equal to zero:

$$\frac{\partial b_1^1(i^*, s, p)}{\partial i} = 1. \quad (14.15)$$

(bi) Suppose that Assumption 2(i) holds. From Lemma 1, this implies $\frac{\partial b_1^1(i^*(0, p), 0, p)}{\partial i} \leq \frac{\partial b_1^1(i^*(1, p), 1, p)}{\partial i}$. For outgroups, at an optimal solution we have $\frac{\partial b_1^1(i^*(0, p), 0, p)}{\partial i} = 1$ for any level of priming, p . But, from Lemma 1, we have

$$1 = \frac{\partial b_1^1(i^*(0, p), 0, p)}{\partial i} < \frac{\partial b_1^1(i^*(1, p), 1, p)}{\partial i}.$$

Thus, $i^*(0, p)$ cannot be the optimal solution for ingroups ($s = 1$), which requires, from part (a), $\frac{\partial b_1^1(i^*(1, p), 1, p)}{\partial i} = 1$. Under the conditions of part (a), b_1^1 is strictly concave. It follows that $i^*(0, p) \leq i^*(1, p)$.

⁴³We need a slightly more nuanced argument for right side and left side limits for the cases $i = 0$ and $i = Y$. But since this is straightforward, and standard, we omit it.

(bii) The proof is as in (bi), hence, it is omitted.

(c) Under the stated conditions, the results in part (bi) and (bii) are true, so (i) $i^*(0, p) \leq i^*(1, p)$ and (ii) $i^*(1, 0) \leq i^*(1, 1)$. Subtracting the first inequality, evaluated at $p = 0$, from the second inequality, we get $i^*(1, 0) - i^*(0, 0) \leq i^*(1, 1) - i^*(1, 0)$. Suppose that we impose the sufficient condition $i^*(1, 0) > i^*(0, 1)$. Then, it follows from the previous inequality that $i^*(1, 0) - i^*(0, 0) \leq i^*(1, 1) - i^*(0, 1)$. ■

Proof of Proposition 6: Using (15.1), and (15.5), we get

$$\frac{\partial W}{\partial z} = -1 + 4\beta(s, p)f'(Z - 4z). \quad (14.16)$$

Since $\frac{\partial^2 W}{\partial z^2} = -16\beta f''(Z - 4z) < 0$, there is a unique maximizer, $z^*(s, p) \leq \frac{1}{4}Z$.

(b) Let $s = 0$, but $p \in \{0, 1\}$. Then, at an interior solution, we have from (15.6) that

$$\frac{\partial W}{\partial z} = -1 + 4\beta(0, p)f'(Z - 4z^*(0, p)) = 0. \quad (14.17)$$

From (15.3)(i), $\beta(0, p) \leq \beta(1, p)$. Thus, it follows from (15.7) that $-1 + 4\beta(1, p)f'(Z - 4z^*(0, p)) \geq 0$. Since we have a unique solution in the domain $Z - z \geq 3z$, and $f'' > 0$, it follows that $z^*(0, p) \leq z^*(1, p)$. Thus, when primed, the dictator gives a higher share to an ingroup receiver.

(c) Proceeding as in (b) and using (i) and (ii) in (15.4), it is straightforward to show that

$$(i) z^*(1, 0) \leq z^*(1, 1), (ii) z^*(0, 1) \leq z^*(0, 0). \blacksquare$$

Proof of Corollary 2: Simply add the two inequalities in Proposition 6c, namely, $z^*(1, 0) \leq z^*(1, 1)$ and $z^*(0, 1) \leq z^*(0, 0)$, and then rearrange to get the required result. ■

15 Supplementary Section

15.1 Identity, priming, and optimal choice in the dictator game: Complete Model

We now provide a more complete model of the dictator game by simultaneously allowing for advantageous inequity (as in the paper) but also disadvantageous inequity. We show that the optimal solution of the dictator is in the domain of advantageous inequity, hence, the results in the paper are unchanged.

The dictator's preferences are given by

$$W(z; s, p) = \begin{cases} (Z - z) - \beta(s, p)f(Z - 4z) & \text{if } Z - z \geq 3z \\ (Z - z) - \alpha(s, p)f(4z - Z) & \text{if } Z - z \leq 3z \end{cases}; 0 \leq \beta \leq \alpha, \beta < 1 \quad (15.1)$$

In (15.1), the dictator derives utility from own material payoffs, $Z - z$; disutility from advantageous payoff differences ($Z - z \geq 3z$) due to altruism (first row), and; disutility from advantageous payoff differences ($Z - z \leq 3z$) due to envy. The parameters of advantageous

and disadvantageous inequity are, respectively, β and α .⁴⁴ When $Z - z = 3z$, the material payoffs of both players are identical, hence, the expressions in both rows coincide and $W = \frac{3}{4}Z$. Self-regarding preferences arise as a special case when $\alpha = \beta = 0$.

The function f which captures disutility from payoff differences is increasing and convex

$$f' > 0, f'' > 0. \quad (15.2)$$

Thus, individuals derive disutility as payoff differences increase and the ‘marginal disutility’ from payoff differences increases at an increasing rate. The empirical evidence shows that typically dictators do not share more than half their endowments, hence, only the first row of (15.1) is empirically relevant; see also Proposition 6 below.

Our assumptions on beliefs made in the trust game are no longer relevant here because the second player is passive. Following Chen and Li (2009) and Dhami et al. (2021), the parameters α and β are influenced by social identity, hence we write them as $\beta(s, p)$ and $\alpha(s, p)$ respectively, where $s, p \in \{0, 1\}$. We make two assumptions below in (15.3) and (15.4).

$$(i) \beta(0, p) \leq \beta(1, p); (ii) \alpha(1, p) \leq \alpha(0, p); p \in \{0, 1\}. \quad (15.3)$$

The evidence shows that people are more altruistic towards ingroups (see (i) in (15.3)), and, from (ii) in (15.3), they feel greater disadvantageous inequity from outgroups (Chen and Li, 2009; Dhami et al., 2021). We extend these insights in the presence of priming as follows.

$$(i) \beta(1, 0) \leq \beta(1, 1), (ii) \beta(0, 1) \leq \beta(0, 0). (iii) \alpha(1, 1) \leq \alpha(1, 0) (iv) \alpha(0, 0) \leq \alpha(0, 1). \quad (15.4)$$

In (15.4), priming heightens the effects of social identity in (15.3). Primed subjects are more altruistic towards ingroups (see (i)) and less altruistic towards outgroups (see (ii)). Primed subjects are less envious of ingroups (see (iii)) and more envious of outgroups (see (iv)).

It will turn out that the conditions (15.3)(ii), (15.4)(iii), and (15.4)(iv) are not relevant because we show in Proposition 6 that, in equilibrium, the dictator does not face disadvantageous inequity. We do not directly test for the conditions in (15.3) and (15.4). However, the preferences in (15.1), in conjunction with the conditions in (15.3) and (15.4) lead to specific predictions (see Proposition 6) that we test with data. A rejection of those predictions is also a potential rejection of (15.3) and (15.4). But we are able to confirm these predictions with the data.

The objective function of the dictator is

$$z^* \in \operatorname{argmax} W(z; s, p), z \in [0, Z], \quad (15.5)$$

where W is defined in (15.1). The objective function is piece-wise differentiable in the domains of advantageous and disadvantageous inequity. We find the optimal value of the function in each domain and then pick the domain where the optimal value is the highest.

Proposition 6. *Consider the optimization problem of the dictator in (15.5)*

(a) *There is a unique solution to the dictator’s problem, $z^*(s, p) \leq \frac{Z}{4}$, for $s, p \in \{0, 1\}$.*

⁴⁴The restriction (i) $\beta < 1$ incorporates evidence which shows that people do not burn their money to try to reduce payoff differences, and (ii) $\beta \leq \alpha$ incorporates evidence that envy is relatively more important than altruism; see Dhami (2019, Vol. II) for the evidence.

- (b) Suppose that $\beta(0, p) \leq \beta(1, p)$, as specified in (15.3), then, $z^*(0, p) \leq z^*(1, p)$, $p \in \{0, 1\}$.
(c) Suppose that $\beta(1, 0) \leq \beta(1, 1)$, and $\beta(0, 1) \leq \beta(0, 0)$, as specified in (15.4), then, (i) $z^*(1, 0) \leq z^*(1, 1)$, (ii) $z^*(0, 1) \leq z^*(0, 0)$.

Proof of Proposition 6: (a) We consider two cases

(a1) (Domain of disadvantageous inequity) Suppose that $Z - z \leq 3z$. Then, from the second row of (15.1), (15.5), and the fact that $f' > 0$ (from (15.2)) we get

$$\frac{\partial W}{\partial z} = -1 - 4\alpha(s, p)f'(4z - Z) < 0.$$

Hence, in the domain of disadvantageous inequity, the optimal value of z , denoted by z^d , is the minimum value that satisfies $Z - z^d \leq 3z^d$, so that $z^d = Z/4$, and the maximum value of the objective function in this domain is $W^d = \frac{3}{4}Z$.

(a2) (Domain of advantageous inequity) Suppose that the dictator's income is relatively greater, i.e., $Z - z \geq 3z$. Then, using the first row of (15.1), and (15.5) we get

$$\frac{\partial W}{\partial z} = -1 + 4\beta(s, p)f'(Z - 4z). \quad (15.6)$$

Since $\frac{\partial^2 W}{\partial z^2} = -16\beta f''(Z - 4z) < 0$, there is a unique maximizer, denoted by z^a . Since this domain allows for the value $z^a = Z/4$, the maximum value in this domain cannot be any less than that in the domain of disadvantageous inequity. It follows that the global maximizer to the problem, $z^*(s, p)$, lies in this domain, so $z^*(s, p) \leq \frac{1}{4}Z$.

(b) First consider the case where $s = 0$, but $p \in \{0, 1\}$. Then, at an interior solution, we have from (15.6) that

$$\frac{\partial W}{\partial z} = -1 + 4\beta(0, p)f'(Z - 4z^*(0, p)) = 0. \quad (15.7)$$

From (15.3)(i), $\beta(0, p) \leq \beta(1, p)$. Thus, it follows from (15.7) that $-1 + 4\beta(1, p)f'(Z - 4z^*(0, p)) \geq 0$. Since we have a unique solution in the domain $Z - z \geq 3z$, and $f'' > 0$, it follows that $z^*(0, p) \leq z^*(1, p)$. Thus, when primed, the dictator gives a weakly higher share to an ingroup receiver.

(c) Proceeding as in (b) and using (i) and (ii) given in (15.4), it is straightforward to show that

$$(i) z^*(1, 0) \leq z^*(1, 1), (ii) z^*(0, 1) \leq z^*(0, 0). \blacksquare$$

From Proposition 6b, dictators share a larger amount with ingroups rather than outgroups for any level of priming. From Proposition 6c, priming increases altruism towards ingroups but decreases altruism towards outgroups. If the data supports the predictions in Proposition 6b,c, then it is also consistent with the assumptions in Proposition 6b,c. The next corollary gives us another testable implication of Proposition 6c in terms of marginal identity effects.

Corollary 2. *Proposition 6c implies $z^*(1, 1) - z^*(0, 1) \geq z^*(1, 0) - z^*(0, 0)$.*

Proof of Corollary 2: Simply add the two inequalities in Proposition 6c, namely, $z^*(1, 0) \leq z^*(1, 1)$ and $z^*(0, 1) \leq z^*(0, 0)$, and then rearrange to get the required result. \blacksquare

15.2 Trustor's first order beliefs

The distributions of Hindu and Muslim trustor's first order beliefs about the trustee's return when the data is pooled for all five investment levels, 1–5, are shown in Figure 4. We show differently shaded histograms for ingroup and outgroup trustees, and the data is pooled together for the primed and the natural treatments. The ingroup/outgroup distributions are significantly different for both Hindu and Muslim trustors (KS test, p -value < 0.05).

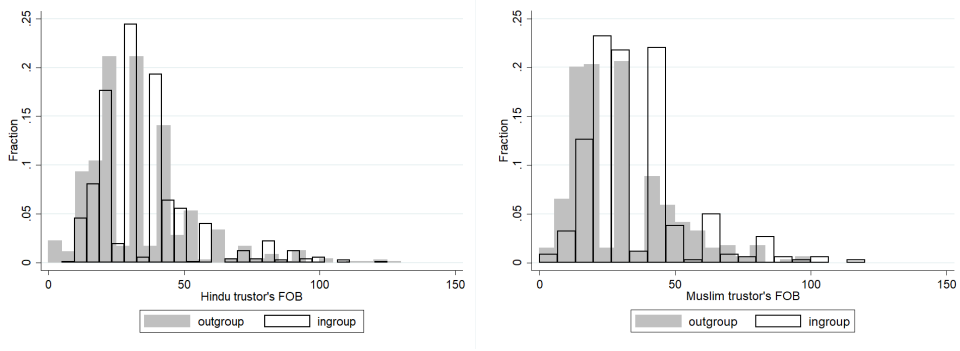


Figure 4: Histograms of Hindu and Muslim trustor's first order beliefs of the trustee's return separated by ingroup and outgroup trustees.

15.3 Trustee's second order beliefs

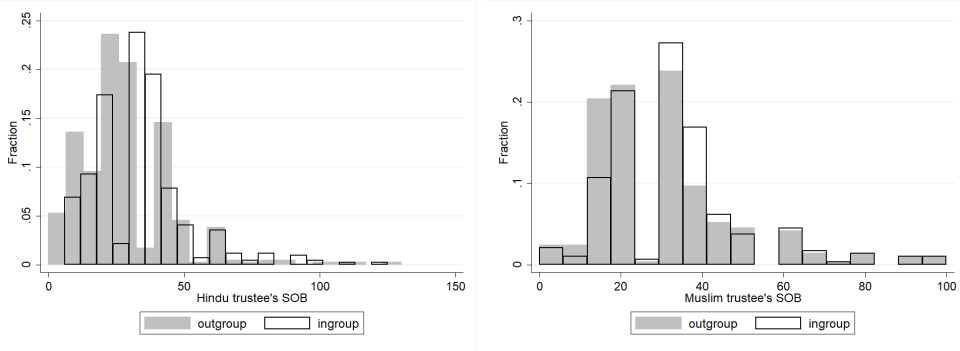


Figure 5: Histograms of Hindu and Muslim trustee's beliefs of the trustor's expected return.

In Figure 5, we pool together the data for the second order beliefs of all trustees at each of the 5 possible investment levels of the trustors, and for both treatments (primed and neutral). We show the distributions separately for Hindu and Muslim trustee's second order beliefs for this pooled data. For both Hindu and Muslim trustees the ingroup/outgroup distributions of expectations of the trustor's beliefs are significantly different (KS test, p -value < 0.05 in both panels).

15.4 Amounts returned by the Trustee

The distributions of the Hindu and Muslim trustees' returns when data are pooled for the five investment levels are shown in Figure 6. Both Hindu and Muslim trustees return higher

amounts to religious ingroup trustors than outgroup trustors. The effect of social identity is significant when the distributions are compared, separately, for subjects of each religion (KS test, p -value < 0.05 in both distributions).

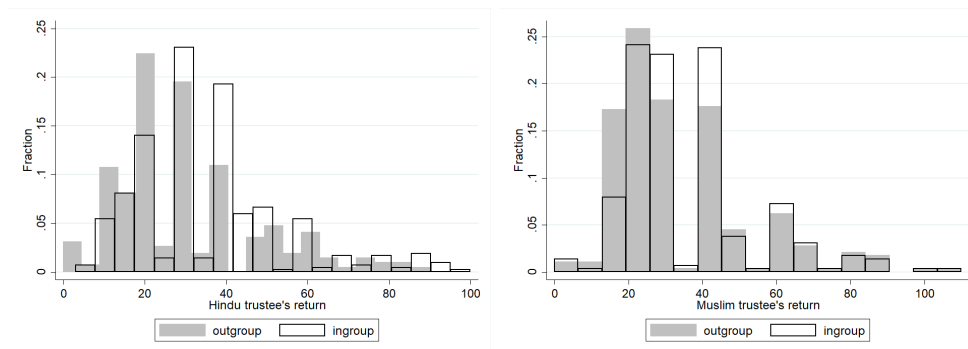


Figure 6: Histograms of Hindu and Muslim trustee's returns. Pooled data for the primed and neutral treatments.

15.5 Amounts invested by trustors

The distributions of Hindu and Muslim trustor's investments, separated by ingroups and outgroups, are shown in Figure 7. For both Hindu and Muslim trustor's, the investment to religious ingroup trustees is higher than to outgroup trustees. For subjects of each religion, the respective ingroup/outgroup effects are significant distributionally (KS test, p -value < 0.05 in both panels).

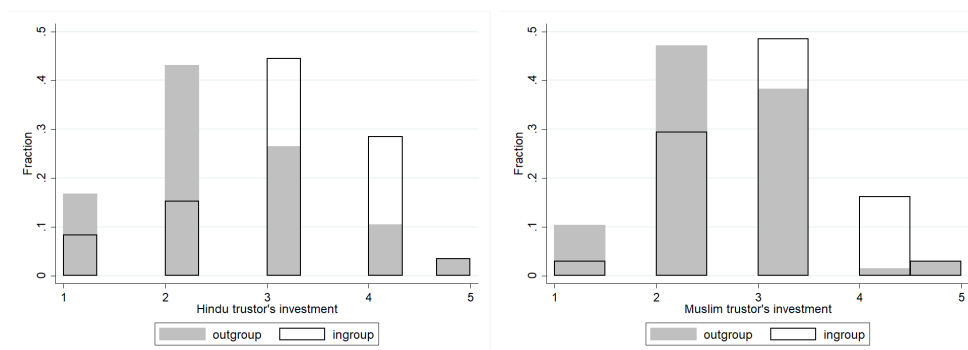


Figure 7: Histograms of Hindu and Muslim trustor's investments separated by ingroups and outgroups.

15.6 Amounts shared by the dictator with the recipient

Figure 8 shows the distributions of transfers made by Hindu and Muslim dictators; the ingroup/outgroup distributions are significantly different (KS test, p -value < 0.05 in both panels).

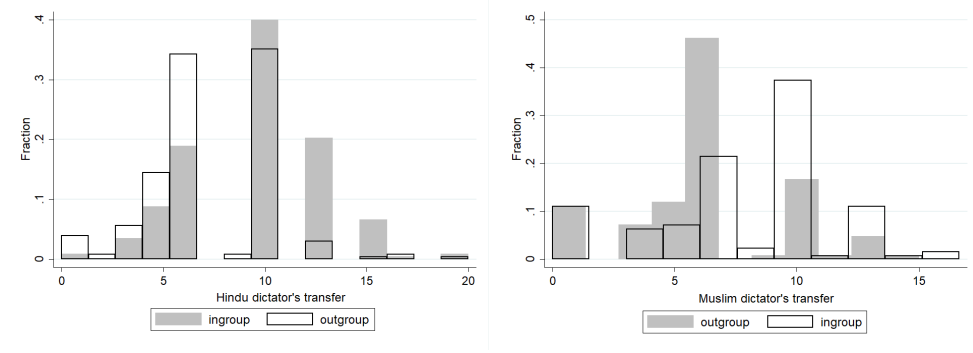


Figure 8: Transfer according to religion in the dictator game.

15.7 No statistical discrimination in trustworthiness

We can also check for statistical discrimination in terms of trustworthiness. The trustee’s decision on the amount that they return to the trustor depends on the amount transferred to them by the trustor. Thus, an ingroup/outgroup comparison can only be established with regard to trustees who have received the same amount of investment from the trustor. Table 16 illustrates the average amounts returned by trustees of different religions when receiving 15, 20, 30, 40, and 45 tokens.

Table 16: Average amount returned by a Hindu and Muslim trustee holding fixed the investment level in the trust game to the levels 15, 20, 30, 40, and 45 tokens.

Trustor’s transfer	15	20	30	40	45
Average return by Hindu	20.67	25.55	34.25	40.16	44.96
Average return by Muslim	19.45	24.78	34.83	42.23	49.83

In order to statistically compare the trustee’s return decision by members of the two religions, we made five comparisons using the Mann-Whitney test, one for each of the amounts received (15, 20, 30, 40, or 45 tokens) by the trustee from the trustor. We strongly confirm ingroup/outgroup effects and trustors of one religion are not uniformly discriminated against by trustees belonging to both religions; this can be checked for all five comparisons at 5% significance level. This result is unchanged if we separated the data in terms of gender or priming.

In sum, there is *no* systematic mistrust of any one religious group in our data from the trust game, either for the trust decision or the trustworthiness decision. Thus, there is no statistical discrimination in the data and we observe significant taste based discrimination.

15.8 Detailed calculations of ingroup outgroup effects for trustees

The following calculations relate to trustees in Models 2 and 4 in Table 10. To keep expressions manageable, we represent regression coefficients that are common to two expressions, and do not affect their differences, by O .

1. Ingroup/outgroup differences for primed Hindu Trustees: Consider the average beliefs of

Hindu trustees ($R = H$) who are primed ($p = 1$) and differences in their SOB about their ingroup trustors ($s = 1$) and outgroup trustors ($s = 0$).⁴⁵

$$E[SOB \mid R = H, p = 1, s = 1] - E[SOB \mid R = H, p = 1, s = 0] = 6.31. \quad (15.8)$$

Thus, primed Hindu trustees believe that Hindu trustors expect 6.31 units more, on average, relative to Muslim trustors. We now consider the remaining cases.

2. Ingroup/outgroup differences for primed Muslim trustees⁴⁶

$$E[SOB \mid R = M, p = 1, s = 1] - E[SOB \mid R = M, p = 1, s = 0] = 4.11. \quad (15.9)$$

Thus, the average ingroup/outgroup differences in SOBs for primed Muslim trustees are positive, but smaller than Hindu trustees.

3. Ingroup/outgroup differences for unprimed Hindu trustees⁴⁷

$$E[SOB \mid R = H, p = 0, s = 1] - E[SOB \mid R = H, p = 0, s = 0] = 3.77. \quad (15.10)$$

Thus, unprimed Hindu trustees believe that Hindu trustors expect 3.77 units more, on average, relative to Muslim trustors.

4. Ingroup/outgroup differences for unprimed Muslim trustees⁴⁸

$$E[SOB \mid R = M, p = 0, s = 1] - E[SOB \mid R = M, p = 0, s = 0] = 0.32. \quad (15.11)$$

Thus, unprimed Muslim trustees believe that Muslim trustors expect 0.32 units more, on average, relative to Hindu trustors. Thus, they expect a lower degree of religious polarization in the behavior of others.

⁴⁵We used the following intermediate calculations.

$$\begin{aligned} E[SOB \mid R = H, p = 1, s = 1] &= -2.57 + 3.77 + 2.54 + O = 3.74 + O. \\ E[SOB \mid R = H, p = 1, s = 0] &= -2.57 + O. \end{aligned}$$

⁴⁶The calculations are:

$$\begin{aligned} E[SOB \mid R = M, p = 1, s = 1] &= -11.44 + 0.32 + 3.79 + O = -7.33 + O. \\ E[SOB \mid R = M, p = 1, s = 0] &= -11.44 + O. \end{aligned}$$

⁴⁷The intermediate calculations are

$$\begin{aligned} E[SOB \mid R = H, p = 0, s = 1] &= 3.77 + O. \\ E[SOB \mid R = H, p = 0, s = 0] &= O. \end{aligned}$$

⁴⁸The intermediate calculations are

$$\begin{aligned} E[SOB \mid R = M, p = 0, s = 1] &= 0.32 + O. \\ E[SOB \mid R = M, p = 0, s = 0] &= O. \end{aligned}$$

15.9 Detailed calculations of determinants of ingroup/outgroup effects for trustees

We use the regression coefficients in Models 2 and 4 in Table 11, in the calculations. We omit regression coefficients that are common to both terms in a difference and denote these other terms with O .

1. Ingroup/outgroup differences for primed Hindu Trustees: We are first interested in the differences in average return, r , of Hindu trustees ($R = H$), between ingroup trustors ($s = 1$) and outgroup trustors ($s = 0$) when they are primed ($p = 1$).⁴⁹

$$E[r | R = H, p = 1, s = 1] - E[r | R = H, p = 1, s = 0] = 2.19. \quad (15.12)$$

Thus, primed Hindu trustees return 2.19 units more, on average, to Hindu trustors relative to Muslim trustors.

2. Ingroup/outgroup differences for primed Muslim trustees⁵⁰

$$E[r | R = M, p = 1, s = 1] - E[r | R = M, p = 1, s = 0] = 3.91. \quad (15.13)$$

The ingroup/outgroup differences in return for primed Muslim trustees are higher than those for Hindu trustees.

3. Ingroup/outgroup differences for unprimed Hindu trustees⁵¹

$$E[r | R = H, p = 0, s = 1] - E[r | R = H, p = 0, s = 0] = 2.80 \quad (15.14)$$

Thus, unprimed Hindu trustees return 2.80 units more, on average, to ingroup trustors.

4. Ingroup/outgroup differences for unprimed Muslim trustees⁵²

$$E[r | R = M, p = 0, s = 1] - E[r | R = M, p = 0, s = 0] = -1.38. \quad (15.15)$$

Thus, unprimed Muslim trustees return 1.38 units less, on average, to ingroup trustors.

⁴⁹We used the following intermediate calculations.

$$E[r | R = H, p = 1, s = 1] = 12.86 + 2.80 - 0.61 - 0.01 - 1.79 + O = 13.25 + O.$$

$$E[r | R = H, p = 1, s = 0] = 12.86 - 0.01 - 1.79 + O = 11.06 + O.$$

⁵⁰The calculations are:

$$E[r | R = M, p = 1, s = 1] = -3.07 - 1.38 + 5.35 - 0.32 - 0.06 + 0.43 + O = 0.95 + O.$$

$$E[r | R = M, p = 1, s = 0] = -3.07 - 0.32 + 0.43 + O = -2.96 + O.$$

⁵¹The intermediate calculations are

$$E[r | R = H, p = 0, s = 1] = 2.80 + O.$$

$$E[r | R = H, p = 0, s = 0] = O.$$

⁵²The intermediate calculations are

$$E[r | R = M, p = 0, s = 1] = -1.38 + O.$$

$$E[r | R = M, p = 0, s = 0] = O.$$

15.10 Detailed calculations on determinants of ingroup/outgroup differences in trustors' first order beliefs

We now consider the findings on the effect of identity and priming on the trustor's FOB based on Models 2, 4 in Table 12.

1. Primed Hindu trustors: We are first interested in the differences in the average first order beliefs of Hindu trustors ($R = H$) about their ingroup trustees ($s = 1$) and outgroup trustees ($s = 0$), when they are primed ($p = 1$).⁵³ We have,

$$E[FOB | R = H, p = 1, s = 1] - E[FOB | R = H, p = 1, s = 0] = 4.94. \quad (15.16)$$

Thus, primed Hindu trustors believe that Hindu trustees will return 4.94 units more, on average, relative to Muslim trustees.

2. Primed Muslim trustors⁵⁴

$$E[FOB | R = M, p = 1, s = 1] - E[FOB | R = M, p = 1, s = 0] = 4.00. \quad (15.17)$$

Thus, the ingroup/outgroup differences in beliefs for Muslim trustors are positive but slightly smaller than Hindu trustors.

3. Unprimed Hindu trustors⁵⁵

$$E[FOB | R = H, p = 0, s = 1] - E[FOB | R = H, p = 0, s = 0] = 3.62. \quad (15.18)$$

Thus, unprimed Hindu trustors believe that Hindu trustees will return 3.62 units more, on average, relative to Muslim trustees.

4. Unprimed Muslim trustors⁵⁶

$$E[FOB | R = M, p = 0, s = 1] - E[FOB | R = M, p = 0, s = 0] = 2.33. \quad (15.19)$$

⁵³We used the following intermediate calculations.

$$E[FOB | R = H, p = 1, s = 1] = 1.33 + 3.62 + 1.32 + O = 6.27 + O.$$

$$E[FOB | R = H, p = 1, s = 0] = 1.33 + O.$$

⁵⁴The calculations are:

$$E[FOB | R = M, p = 1, s = 1] = -5.52 + 2.33 + 1.67 + O = -1.52 + O.$$

$$E[FOB | R = M, p = 1, s = 0] = -5.52 + O.$$

⁵⁵The intermediate calculations are

$$E[FOB | R = H, p = 0, s = 1] = 3.62 + O.$$

$$E[FOB | R = H, p = 0, s = 0] = O.$$

⁵⁶The intermediate calculations are

$$E[FOB | R = M, p = 0, s = 1] = 2.33 + O.$$

$$E[FOB | R = M, p = 0, s = 0] = 0.$$

Thus, unprimed Muslim trustors believe that Muslim trustees will return 2.33 units more, on average, relative to Hindu trustees.

15.11 Detailed calculations on ingroup/outgroup differences for dictators

The following calculations are based on Models 2, 4 in Table 15.

1. Primed Hindu dictators: We are first interested in the differences in transfers, t , of Hindu dictators ($R = H$) between their ingroup ($s = 1$) and outgroup ($s = 0$) receivers, when they are primed ($p = 1$).⁵⁷

$$E[t \mid R = H, p = 1, s = 1] - E[t \mid R = H, p = 1, s = 0] = 2.52. \quad (15.20)$$

Thus primed Hindu dictators gave 2.52 units more, on average, to Hindu receivers relative to Muslim receivers.

2. Primed Muslim dictators⁵⁸

$$E[t \mid R = M, p = 1, s = 1] - E[t \mid R = M, p = 1, s = 0] = 1.83. \quad (15.21)$$

Thus, primed Muslim dictators give 1.20 units more to their ingroup receivers relative to the outgroup receivers.

3. Unprimed Hindu dictators⁵⁹

$$E[t \mid R = H, p = 0, s = 1] - E[t \mid R = H, p = 0, s = 0] = 1.56 \quad (15.22)$$

Thus, unprimed Hindu dictators give 1.56 units more, on average, to Hindu receivers relative to Muslim receivers.

4. Unprimed Muslim dictators⁶⁰

$$E[t \mid R = M, p = 0, s = 1] - E[t \mid R = M, p = 0, s = 0] = 0.58. \quad (15.23)$$

⁵⁷We used the following intermediate calculations.

$$E[t \mid R = H, p = 1, s = 1] = -0.58 + 1.56 + 0.96 + O = 1.94 + O.$$

$$E[t \mid R = H, p = 1, s = 0] = -0.58 + O.$$

⁵⁸The calculations are:

$$E[t \mid R = M, p = 1, s = 1] = -0.05 + 0.58 + 1.25 + O = 1.78 + O.$$

$$E[t \mid R = M, p = 1, s = 0] = -0.05 + O.$$

⁵⁹The intermediate calculations are

$$E[t \mid R = H, p = 0, s = 1] = 1.56 + O.$$

$$E[t \mid R = H, p = 0, s = 0] = O.$$

⁶⁰The intermediate calculations are

$$E[t \mid R = M, p = 0, s = 1] = 0.58 + O.$$

$$E[t \mid R = M, p = 0, s = 0] = O.$$

Thus, unprimed Muslim dictators also give 0.58 units more, on average, to Muslim receivers relative to Hindu receivers.

References

- [1] Adida, Claire L. and Laitin, David and Valfort, Marie-Anne, (2016) One Muslim is Enough! - Evidence from a Field Experiment in France. *Annals of Economics and Statistics* 121/122: 121-160.
- [2] Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics* 85: 207-34.
- [3] Algan, Y., and P. Cahuc (2014) Trust, Growth, and Well-Being: New Evidence and Policy Implications. Chapter 2 in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 2 of *Handbook of Economic Growth*, pp. 49 – 120. Elsevier.
- [4] Anderson, L.R., and Mellor, J.M. (2009). Religion and cooperation in a public goods experiment. *Economic Letters* 105(1): 58–60.
- [5] Anderson, L., Mellor, J., Milyo, J. (2010). Did the devil make them do it? The effects of religion in public goods and trust games. *Kyklos* 63: 163–175.
- [6] Arrow, K. J. (1972). Gifts and Exchanges. *Philosophy and Public Affairs* 1(4), 343-362.
- [7] Arrow, K. J. (1973) The Theory of Discrimination. in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press. pp. 3–33.
- [8] Arrow, K. J. (1998) What Has Economics to Say about Racial Discrimination? *Journal of Economic Perspectives*, 12: 91–100.
- [9] Arruñada, B.. (2010) Protestants and Catholics: Similar Work Ethic, Different Social Ethic. *Economic Journal*. 120(547) :890–918.
- [10] Battigalli, P. and Dufwenberg, M. (2009). Dynamic Psychological Games. *Journal of Economic Theory*, 144(1): 1–35.
- [11] Battigalli, P. and Dufwenberg, M. (2022). Belief-Dependent Motivations and Psychological Game Theory. *Journal of Economic Literature* 60(3):833–82.
- [12] Bellemare, C., Sebald, A. and Strobel, M. (2011). Measuring the Willingness to Pay to Avoid Guilt: Estimation Using Equilibrium and Stated Belief Models. *Journal of Applied Econometrics*, 26(3): 437–453.
- [13] Benjamin, D. J ., Choi, J. J., and Fisher, G. (2016) Religious identity and economic behavior. *Review of Economics and Statistics*. 98(4): 617-637.
- [14] Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10: 122–142.

- [15] Brañas-Garza P, Espín AM, Neuman S (2014) Religious Pro-Sociality? Experimental Evidence from a Sample of 766 Spaniards. *PLoS ONE* 9(8): e104685.
- [16] Chakravarty, Surajeet, Miguel A. Fonseca, Sudeep Ghosh, and Sugata Marjit. (2016). Religious Fragmentation, Social Identity and Cooperation: Evidence from an Artefactual Field Experiment in India. *European Economic Review* 90(C): 265–79.
- [17] Chen, Y., and Li, S. (2009) Social Identity and Group Preferences. *American Economic Review* 99(1) 431-457.
- [18] Chuah, S.H., Fahoum, R., Homann, R (2013). Fractionalization and trust in India: a field-experiment. *Economics Letters* 119(2): 191-194.
- [19] Chuah, S. H., Gächter, S., Hoffmann, R., Tan, J. H. W. (2016) Religion, discrimination and trust across three cultures. *European Economic Review* 90: 280–301.
- [20] Chuah, S.H., Hoffmann, R., Ramasamy, B., Tan, J.H.W. (2014). Religion, ethnicity and cooperation: an experimental study. *Journal of Economic Psychology* 45: 33–43.
- [21] Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press: Oxford.
- [22] Dhami, S. (2019). *The Foundations of Behavioral Economic Analysis. Volume II: Other-Regarding Preferences*, Oxford University Press: Oxford.
- [23] Dhami, S. (2020). *The Foundations of Behavioral Economic Analysis. Volume IV: Behavioral Game Theory*, Oxford University Press: Oxford.
- [24] Dhami, S. (2020). *The Foundations of Behavioral Economic Analysis. Volume V: Bounded Rationality*, Oxford University Press: Oxford.
- [25] Dhami, S., Arshad, J., and al-Nowaihi, A. (2022). Psychological and social motivations in microfinance contracts: Theory and evidence. *Journal of Development Economics* 158: 102912
- [26] Dhami, S., Manifold, E., and al-Nowaihi, A. (2021) Identity and Redistribution: Theory and Evidence. *Economica*, 88(350): 499-531.
- [27] Dhami, S., Wei, M. and al-Nowaihi, A. (2019). Public Goods Games and Psychological Utility: Theory and Evidence. *Journal of Economic Behavior and Organization*. 167: 361-390.
- [28] Dufwenberg, M., and Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*. 47(2): 268-98.
- [29] Dufwenberg, M., and Kirchsteiger, G. (2019). Modelling Kindness. *Journal of Economic Behavior & Organization* 167, 228-34.

- [30] Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *Quarterly Journal of Economics* 112: 1203-50.
- [31] Fehr, E., Fischbacher, U., Rosenbladt, B.V., Schupp, J., Wagner, G.G. (2003). A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioural experiments into representative survey. IZA Discussion Paper No. 715.
- [32] Fehr, E., and Schmidt, K. M. (1999). A Theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*. 114 (3): 817-68.
- [33] Fershtman, C., Gneezy, U. (2001). Discrimination in a segmented society: an experimental approach. *Quarterly Journal of Economics*. 116(1):351–377.
- [34] Fukuyama, Francis (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press: New York.
- [35] Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior* 1(1): 60–79.
- [36] Gupta, G., Mahmud, M., Maitra, P., Mitra, S., and Neelim, A. (2018) Religion, minority status and trust: Evidence from a field experiment. *Journal of Economic Behavior & Organization* 146: 180-205.
- [37] Golder, M. (2016) Far Right Parties in Europe. *Annual Review of Political Science*. 19: 477–97.
- [38] Guiso, L., Sapienza, P., and Zingales, L. (2003). People's Opium? Religion and Economic Attitudes. *Journal of Monetary Economics* 50(1): 225-82.
- [39] Johansson-Stenman, O., Mahmud, M., Martinsson, P., 2009. Trust and religion: experimental evidence from rural Bangladesh. *Economica* 76, 462–485.
- [40] Johnson, N. D. and Mislin, A. A., 2011. Trust games: a meta-analysis. *Journal of Economic Psychology* 32(5): 865–889.
- [41] Karlan, D.S (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*. 95(5): 1688–1699.
- [42] Khalmetski, K., Ockenfels, A. and Werner, P. (2015). Surprising Gifts: Theory and Laboratory Evidence. *Journal of Economic Theory*, 159: 163–208.
- [43] Knack, S. and Keefer, K. (1997) Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics* 112(4): 1251-1288
- [44] Lane, T. (2016). Discrimination in the Laboratory: A Meta- Analysis of Economics Experiments. *European Economic Review* 90(C): 375–402.

- [45] Mantilla, C., Zhou, L., Wang, C., Yang, D., Shen, S., and Seabright, P. (2021) Favoring your in-group can harm both them and you: Ethnicity and public goods provision in China. *Journal of Economic Behavior and Organization* 185: 211–233
- [46] Mauersberger, F. and Nagel, R. (2018). Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework, in *Handbook of Computational Economics*. Vol. 4, Elsevier, pp. 541–634.
- [47] Pollard, R. (2022) Anti-Muslim Hatred Is Shaping India’s Elections. Bloomberg UK.
- [48] Putnam, R. (1993) *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press: Princeton, NJ.
- [49] Rabin M. (1993) Incorporating fairness into game theory and economics. *The American economic review*. 83(5): 1281-1302.
- [50] Rohrbaugh, J. and Jessor, R.(1975).Religiosity in youth: a personal control against deviant behavior. *Journal of Personality* 43(1): 136–155.
- [51] Sahoo, N. (2020) *Mounting Majoritarianism and Political Polarization in India*. Carnegie Endowment for International Peace: Washington DC.
- [52] Shariff, A. F. and Norenzayan, A. (2007) God Is Watching You: Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game. *Psychological Science*. 18: 803–809.
- [53] Shariff, A. F., Willard, A. K., Andersen, T., and Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, 20(1): 27-48.
- [54] Tan, J.H.W. and Vogel, C. (2008). Religion and trust: an experimental study. *Journal of Economic Psychology* 29: 832–848.
- [55] Traunmüller, R. (2011). Moral Communities? Religion as a Source of Social Trust in a Multilevel Analysis of 97 German Regions. *European Sociological Review*, 27(3), pp. 346-363.
- [56] Vaishnav, M. (2019, eds.) *The BJP in Power: Indian Democracy and Religious Nationalism*. Carnegie Endowment for International Peace: Washington DC.
- [57] Xia, W., Guo, X., Luo, J., Ye, H., Chen, Y., Chen, S., and Xia, W., (2021). Religious identity, between-group effects and prosocial behavior: Evidence from a field experiment in China, *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)*, 91(C): 101665.
- [58] Zak, P. J. and Knack, S. (2001). Trust and growth. *Economic Journal*, 111: 295-321.