NBER WORKING PAPER SERIES

# LEARNING ABOUT OUTGROUPS: THE IMPACT OF BROAD VERSUS DEEP INTERACTIONS

Anujit Chakraborty
Arkadev Ghosh
Matt Lowe
Gareth Nellis

Working Paper 35884
http://www.nber.org/papers/w35884

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2026

This project was pre-registered in the AEA RCT Registry (AEARCTR-0013126), and approved by ethics boards at UBC (H23-03476), Duke (2024-0213), UCSD (809306), and UC-Davis (2136225-1). We are grateful for funding from the Weiss Fund for Research in Development Economics, UCSD (Senate General Research), Duke (Social Science Research Institute), CEGA (Development Economics Challenge), and the Institute for Humane Studies. We thank Diptorup Chatterjee, Aura Gonzalez, Meghna Sinha Ray, Jingjing Zhang, and Ved Sharma for outstanding research assistance. For comments and advice, we especially thank Siwan Anderson, Sid Baral, Karen Ferree, Patrick Francois, Sridipta Ghatak, and Gaurav Khanna. Finally, we are deeply grateful to our enumerators, factory managers, and survey team without whom none of this research would have been possible. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Anujit Chakraborty, Arkadev Ghosh, Matt Lowe, and Gareth Nellis. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Learning About Outgroups: The Impact of Broad Versus Deep Interactions
Anujit Chakraborty, Arkadev Ghosh, Matt Lowe, and Gareth Nellis
NBER Working Paper No. 35884
October 2026
JEL No. C93, D83, D9, Z13

## ABSTRACT

We hypothesize that *broad* contact, involving brief interactions with multiple outgroup members, and *deep* contact, meaning longer interactions with a single outgroup member, play distinct roles in shaping intergroup relations. We set up a factory in India and recruited Hindu and Muslim men to work in pairs on joint production tasks. We randomly assigned participants to work either with the same ingroup or outgroup partner daily (deep contact), a different outgroup partner each day (broad contact), or to a control group. Deep contact strengthens social and economic ties with the outgroup partner. But only broad contact reduces the majority group's misperceptions about outgroup strangers. Using within-treatment variation in partners' pre-treatment productivity, we show that participants in both arms learn about their partners, but broad participants also generalize these belief updates to strangers. These findings align with a model in which independent sampling promotes learning about outgroups more than prolonged interaction with a single individual does. Nevertheless, neither type of contact changes behavior toward the wider outgroup.

Anujit Chakraborty
University of California, Davis
Department of Economics
chakraborty@ucdavis.edu

Arkadev Ghosh
Duke University
Economics
and NBER
arkadev.ghosh@duke.edu

Matt Lowe
University of British Columbia
matt.lowe@ubc.ca

Gareth Nellis
University of Pennsylvania
gnellis@sas.upenn.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/13126

## 1 Introduction

People often hold false or exaggerated beliefs about outgroups (Bordalo et al., 2016; Bursztyn and Yang, 2022). They regard members of other races, religions, and ethnicities as greedy, untrustworthy, backward, unclean, or aggressive.[1] Collaborative contact has emerged as a leading tool for reducing such enmity (Allport, 1954; Mousa, 2020; Lowe, 2021; Bursztyn et al., 2024). By disrupting the homophily of social networks (Jackson et al., 2023), the hope is that cross-group interactions will build "bridging" social capital, paving the way for more cohesive communities (Putnam, 2000; Varshney, 2001). But despite early optimism, recent experiments paint a mixed picture (e.g., Scacco and Warren, 2018; Elwert et al., 2023; Grady et al., 2023; Zhou and Lyall, 2025). A meta-analysis of pre-registered trials finds that intergroup contact has only modest positive impacts on prejudice, with improvements in attitudes toward outgroup members directly interacted with typically not carrying over to attitudes toward the outgroup as a whole (Lowe, 2025). What is it that prevents the effects of intergroup contact from generalizing, and how can intergroup contact be optimally structured to correct misperceptions about outgroups?

This paper advances a new hypothesis concerning generalization. When people have favorable interactions with only a few outgroup members, they might view those individuals as exceptional rather than prototypical of the outgroup at large (Hewstone and Brown, 1986; Brown et al., 1999; Paluck and Clark, 2020), perpetuating inaccurate beliefs about that group. Interactions with a greater number of outgroup members may help disrupt this logic of outgroup exceptions, and thereby generate more generalized belief updating. We call this the *sampling channel* of contact. If independent sampling is what promotes learning, then *broad* contact—short interactions with many outgroup members—may do more to undermine prejudice than *deep* contact, meaning many interactions with one outgroup member.

Although this claim has intuitive appeal, it is not self-evidently true. Prominent work in psychology takes the opposite track. It argues that for contact to successfully reduce distrust and hatred, the potential for friendship must exist, with intimate interactions being more effective in mitigating prejudice than casual, surface-level encounters (Amir, 1976). Indeed,

[1] For example, Roma communities in Europe have long been vilified as "thieves" (Bracic, 2020). In Southern Sudan, the Nuer are considered "warlike ... intractable and fierce" by neighboring tribes (Johnson, 1981, 512). Orientalist stereotypes were central to justifications of Western imperialism (Said, 1978). This lack of knowledge, and the distorted imaginations and misrepresentations that result, may have severe consequences. Ethnic diversity is associated with lower public goods provision (Alesina and La Ferrara, 2005), poor productivity (Hjort, 2014), labor market discrimination (Banerjee et al., 2009), and sometimes violence (Horowitz, 2001; Collier and Hoeffler, 2004).

according to Pettigrew (1998, 76), a fifth condition for effective contact—complementing the four conditions of Allport (1954)—is the opportunity for participants to establish meaningful friendships: "Such opportunity implies close interaction that would make self-disclosure and other friendship-developing mechanisms possible." We refer to this as the *friendship channel*, and it predicts the superiority of deep contact.

Real-world contact interventions fall along the full range of the broad-deep spectrum: for example, the random assignment of roommates is a common example of deep contact (Boisjoly et al. 2006; Camargo et al. 2010; Corno et al. 2022), while the random assignment of volunteers and teachers to different regions of a country leads to broader interactions (Okunogbe 2023; Bagues and Roth 2023; Friedman et al. 2024). More generally, economic forces lead to deep interactions in some settings, and broad interactions in others. For example, weak contract enforcement leads to the deep interactions of relational contracting (Macchiavello 2022), while the decline of intensive kinship norms leads to broad interactions with a wider network (Ghosh et al. 2023). With total contact time fixed, deeper interaction encourages investment in particular relationships, while broader interaction provides more independent observations of the group. We study the implications of this allocation trade-off for intergroup relations.

We propose a simple model to formalize how contact with outgroup members shapes belief updating, distinguishing between two processes: updating beliefs about the individual interacted with, and generalizing those updates to the outgroup as a whole. The model makes three predictions. First, the returns to contact are concave: individuals with more interaction with the outgroup (e.g., minority-group members who tend to have higher frequency contact with majority-group members) update their beliefs less with additional exposure. Second, the model reveals a *sampling trade-off*: broad contact—which offers a larger number of independent signals—enables generalized updating about the outgroup, even as each episode of shorter-duration contact limits learning about any particular outgroup individual interacted with. Deep contact provides more detailed information about individual partners, but limits generalized updating. Third, to interpret heterogeneity across the attributes we measure, we use an extension of the baseline model that distinguishes between highly visible attributes of outgroups, like productivity at work, versus less visible attributes, like friendship potential. The model predicts that less visible attributes, which are learned through forming friendships, are inherently hard to generalize. This is because the cost of friendship formation is only justified when the friendship pays off over a sufficiently long period of interaction, which, along with the sampling trade-off inherent in such deep contact, reduces the scope for

generalized updating. Therefore, even though friendship formation should be higher under deep contact, any belief updating about less visible attributes would not generalize towards the larger outgroup unless the total contact hours are expanded to enable an intervention that is both deep and broad.

We ran a field experiment with 964 young adult men in India to study the effects of broad versus deep contact. Our study site is in a district in West Bengal that lies on the India-Bangladesh border, where tensions between Hindus (the majority group) and Muslims (the minority group) have erupted into violence in recent years. We set up a factory in which participants made paper bags in pairs for six days. The task required constant physical interaction and coordination between partners. For the main randomization, we assigned men to work for six days either with the same outgroup partner every day (*Outgroup, deep*), or with a different outgroup partner each day (*Outgroup, broad*). In a third arm, Hindus worked for six days with the same ingroup partner (*Ingroup, deep*). A fourth arm including both Hindus and Muslims served as a pure control condition, only responding to surveys. We took daily measures of productivity and affective mood in the workplace, where attendance was 96%. We fielded an endline survey two to three weeks after the intervention ended. The endline measured beliefs, with participants making incentivized guesses about the preferences and behaviors of unknown outgroup members from a holdout sample, along with social and economic ties, willingness to interact with outgroup members, and altruism.

We find that broad and deep contact lead to different types of partner interactions, which in turn produce endline differences in behaviors, beliefs, and the degree to which they generalize. Worker morale is higher in the deep condition than in the broad condition. During the intervention, subjects assigned to the *Outgroup, broad* condition are more critical of their own work and their partner's performance than subjects assigned to the *Outgroup, deep* condition. They discuss fewer topics overall than their deep counterparts, and perhaps as a result, broad subjects report being less happy ($0.09\sigma$, $p = 0.06$) and less friendly with their work partners ($0.47\sigma$, $p < 0.01$) than deep subjects. These differences appear even on the first day, showing that the mere anticipation of continued interaction with a partner promotes happiness and friendship.[2]

The differences in workplace atmosphere translate into different social and economic ties after the work ends. At endline, 60% of *Outgroup, deep* subjects stated that they had

[2]The lower camaraderie seen in the broad condition resonates with work in organizational behavior, which highlights associations between high employee turnover, low staff morale, and an absence of close friendships at work (e.g., Harter et al., 2002).

had contact with their assigned work partner over the previous week, which, remarkably, is almost the same percentage seen among *Ingroup, deep* subjects. In contrast, subjects in *Outgroup, broad* are 14 percentage points less likely than *Outgroup, deep* subjects to have had contact with their partners ($p < 0.01$). This deep versus broad difference extends to economic interactions. *Outgroup, broad* participants are three percentage points less likely to have borrowed from or lent money to any work partner ($p = 0.03$). Focusing on the level of the partner pair, we find that those in *Outgroup, broad* are significantly less likely to have had in-person or phone contact with one another, and are also 11 percentage points less likely to have received help finding paid work from their partner ($p < 0.01$). Therefore, deep contact in a work environment like ours is more conducive to stimulating intergroup social and economic ties among work-partners.

In contrast, we find that only broad contact causes belief updating that generalizes to the outgroup as a whole. Hindu subjects assigned to the *Outgroup, broad* condition make more accurate guesses about the characteristics of unknown Muslims in the holdout sample vis-à-vis *Outgroup, deep* ($0.09\sigma$, $p = 0.03$). These effects are driven by guesses about outgroup attributes that were straightforward to observe in the factory setting, what we call visible attributes in our model (though we caveat that we did not pre-register our classification of attributes by visibility): the outgroup's average productivity (seen through bag-making), religious identity attachments (seen through clothing, the wearing of religious symbols, and praying behaviors), and family-size preferences (with family being a ubiquitous conversation topic in Bengali society). Meanwhile, assignment to the *Outgroup, broad* condition does nothing to improve Muslims' prediction accuracy about unknown Hindus. This null finding aligns with our theoretical result that contact exhibits diminishing returns. In fact, we demonstrate empirically that Muslims, as a minority, are much more exposed to Hindus in their everyday lives than vice versa; consistent with this, Muslims hold more accurate beliefs about the outgroup to begin with, leaving less scope for updating from additional contact. In a placebo test, we find that neither of the outgroup conditions increases prediction accuracy about ingroups, as we should expect.

While the two types of contact differ in their effects on holdout-sample belief accuracy, consistent with the model's predictions, both types of contact cause participants to report more similar guesses about ingroup and outgroup strangers relative to the control group. It appears that contact of any sort engenders recognition of a common humanity—countering the tendency to view outgroups as fundamentally alien from ingroups.

To test the sampling channel more directly, we turn to a domain for which we have proxies of the signals observed by participants: productivity. We ask whether the *same* change in encountered productivity generates more generalized updating under broad contact. Exploiting random variation in assigned partners' pre-treatment productivity, we find a stronger response of outgroup beliefs relative to ingroup beliefs under broad than deep contact ($p < 0.01$). Estimated gains in predicting partners' own productivity are similar across arms, suggesting that the breadth of contact is shaping belief generalization specifically, and not just learning of all kinds.

While broad beats deep at generalized belief updating, our theory predicts that deep may beat broad at belief updating about the work partner. With the caveat of more limited data on beliefs about partners, we see little support for this prediction, whether considering productivity (as discussed above), or other belief dimensions. Both broad and deep contact cause belief updating about the partner, but the effects cannot be statistically distinguished from each other. This suggests that little additional learning about partners occurs after the first day, at least for the belief dimensions we measure. The only exception is in learning about "friendship potential," revealed by stronger ties under deep than broad. This aligns with our theory that deeper interactions allow greater scope for uncovering such potential, reflected in our data through richer conversations during the work period and continued support afterward (e.g., borrowing, lending, or job search help). As discussed earlier, a likely mechanism is that deeper interactions strengthen incentives to invest in friendship and to learn traits needed to sustain it.

Although our results offer support for the model's predictions around generalized belief updating, neither type of contact leads to generalized behavior change: we find no effects on donations to an outgroup stranger in a dictator game, incentivized willingness to work with an outgroup stranger, the number of close outgroup friends, and a WhatsApp-based "audit" measure of civility toward an outgroup stranger. Future research might explore whether the mechanisms behind generalization differ for beliefs versus behavior.

Finally, despite negative effects of broad contact on subjective assessments of productivity, we estimate null effects on objectively measured productivity. One possible explanation for the null is the simple production technology, which facilitates the rapid updating about (outgroup) partner ability that we observe, which limits productivity losses due to misperceptions (Ghosh, 2025). In turn, the negative subjective assessments may be a spillover effect from weaker interpersonal relations, as we also found in a subsequent experiment with cross-caste partnerships (Chakraborty et al., 2026).

Stepping back, our results prompt questions about external validity. Is the lack of generalization with deep contact itself a generalizable finding, or due to our six days of contact not being deep enough? We make two points here. First, we show that our experiment's breadth and depth of contact fall within the range of previous contact interventions, although at the upper end (the 92nd and 84th percentiles). This suggests that our design choices are in the right range to apply to typical intergroup contact programming. Second, we see signs that our deep contact is fairly deep in practice. In particular, we find little difference between *Outgroup, deep* and *Ingroup, deep* participants in their reported interaction and closeness with partners at endline. Our results then speak both to plausible policy decisions (around whether one week's contact should prioritize breadth or depth), and to a more conceptual understanding of what deep, meaningful, intergroup contact can achieve. Nevertheless, we acknowledge that our experiment serves as a proof of concept, and that deeper and broader contact interventions deserve future study.

Our overall message, summarized in Table 1, reflects the two central mechanisms in our conceptual framework. Deep contact is more effective at building relationships with particular outgroup partners, consistent with greater relational investment when future interaction is anticipated: participants form stronger social ties during the intervention, and these ties persist after the study concludes. Broad contact is more effective at improving beliefs about the outgroup at large, consistent with the sampling mechanism: exposure to more independent outgroup members provides a better basis for generalizing to the wider group.

Should our results be greeted as optimistic or pessimistic by proponents of intergroup contact? On the one hand, broad contact helps erase some false beliefs about outgroups. Such beliefs can be pernicious. More pessimistically, however, our results underscore the challenges to engineering generalized behavioral change via contact alone. There is a catch-22. Expansive sampling (through broad interaction) is informative about outgroups, but generates weaker intergroup ties; friendship (through deep interactions) offers stronger ties, yet the sample size is too small for a rational decision-maker to draw general conclusions. In practice, creating contact that is both broad *and* deep—the theoretical ideal—may be very costly.

Our study contributes to various literatures. A large literature on social learning studies how individuals learn *from* others, and how the structure of social networks determines information transmission (Golub and Jackson 2012; Banerjee et al. 2013; Mobius and Rosenblat 2014; Mobius et al. 2015; Chandrasekhar et al. 2020; Banerjee et al. 2021; Conlon et al. 2022). We instead study how the structure of repeated interactions shapes what

Table 1: Summary Effects of *Outgroup, broad* Versus *Outgroup, deep*

| Outcome | Variable | Winner | Relevant mechanism |
|---|---|---|---|
| **Panel A: During work** | | | |
| Social relations | 1. Happiness | OG Deep | Relational investment |
| | 2. Partner friendship | OG Deep | Relational investment |
| | 3. Number of conversation topics | OG Deep | Relational investment |
| | 4. Happiness, friendship (Day 1) | OG Deep | Anticipated interaction |
| Productivity | 1. Quantity | None | — |
| | 2. Quantity × Quality | None | — |
| **Panel B: After work ends** | | | |
| Social relations | 1. Sustained social ties | OG Deep | Relational investment |
| | 2. Partner helped finding work or with financial assistance | OG Deep | Relational investment |
| Beliefs about outgroup strangers | 1. Overall accuracy | None | Sampling trade-off and diminishing returns in belief updating |
| | 2. Accuracy (readily observed attributes) | OG Broad | |
| | 3. Accuracy (Hindus only) | OG Broad | |
| | 4. Accuracy (Muslims only) | None | |
| | 5. Tracks partners' productivity | OG Broad | Sampling trade-off |
| Beliefs about partner | 1. Accuracy | None | Partner learning |
| Intergroup behaviors | 1. Towards outgroup strangers | None | — |

*Notes:* This table summarizes outcome comparisons between *Outgroup, deep* (OG Deep) and *Outgroup, broad* (OG Broad). **Winner** denotes the treatment that produced a statistically significantly more positive effect on a given outcome. "None" indicates no statistically significant difference between the two. The final column indicates the component of the conceptual framework most directly related to each outcome; an em dash indicates that the framework does not make a prediction.

individuals learn *about* others, and the groups from which those others are drawn. In doing so, we contribute to empirical work on intergroup contact. Contact experiments typically find positive effects of the four scope conditions posited by Allport (1954): common goals (Lowe, 2021), intergroup cooperation (Ghosh, 2025), equal status (Greene et al., 2025), and the support of authorities (Chaudhry et al., 2025). Nevertheless, recent meta-analytic work finds that even contact that satisfies all four Allport conditions has limited generalized effects (Lowe, 2025), despite generalization being at the heart of the contact hypothesis—interaction with outgroup members should reduce prejudice toward outgroup strangers. We make progress on the problem of non-generalization by introducing an overlooked dimension of contact, namely, whether interactions are deep or broad in nature. We conclude that the sampling channel promotes belief generalization, while the friendship channel emphasized by psychologists promotes tie creation (Pettigrew 1998; Davies et al. 2011).[3]

Beyond beliefs about outgroups, many papers document misperceptions about others,

[3]Psychologists have long theorized about what dimensions of contact are specifically likely to maximize *generalization*, though without a focus on the breadth and depth of contact. For a discussion, see Tropp et al. (2024).

whether outgroup members or not (reviewed in Bursztyn and Yang (2022)), and some papers show that providing objective information to correct misperceptions can change subsequent behavior (Bursztyn et al. 2020). Few papers shed light on the origins and evolution of such misperceptions. One example is Matavelli (2024), who finds that discussions with peers about masculinity quickly reduce misperceptions around peers' opinions on masculinity norms, while Webb (2024) shows that the observation of ingroup behavior shifts perceptions about the ingroup. We instead show how the structure of outgroup contact determines the evolution of misperceptions, and find that broad contact makes generalized beliefs more responsive to partners' productivity.

Finally, our conceptual framework builds on theoretical work on stereotypes. While prior theory often focuses on explaining why stereotypes arise, for example through statistical discrimination under asymmetric information (Phelps, 1972; Arrow, 1973), coarse categorization of encounters with the outgroup (Fryer and Jackson, 2008), selective (in)attention to context (Schwartzstein, 2014), or cognitive heuristics (Fryer and Jackson, 2008; Bordalo et al., 2016), we model how stereotypes evolve through repeated interactions with outgroup members. Our focus on the dynamics of *inaccurate* stereotypical beliefs differentiates our approach from the statistical discrimination literature, which assumes beliefs are *accurate* on average. Instead, our model aligns more closely with recent empirical research on inaccurate statistical discrimination, which shows that stereotypes can be flawed and subject to change when confronted with new information (Bohren et al., 2019, 2025). However, while this literature typically focuses on how stereotypes drive individual-level discrimination, our primary outcome of interest is the tradeoff between broad versus deep interactions and the extent of generalization in updating. The decision-makers in our model are behavioral only in that they begin with a systematically inaccurate prior, but they rationally update their beliefs using Bayes' rule when confronted with contradictory evidence. We demonstrate that this simple almost fully Bayesian model can broadly explain our empirical patterns.

## 2 Context, intervention, and experiment design

### 2.1 Local social relations

We conducted our study in Barasat, a city in North 24 Parganas district of West Bengal, eastern India. We recruited participants both from the city itself and from nearby village clusters, focusing on low-income neighborhoods. The district is home to both Hindu (76%)

and Muslim (24%) communities and is a microcosm of the state's broader cultural and religious diversity. Communal relations have become increasingly strained in recent years. Religious festivals like Durga Puja and Ram Navami have become flashpoints, inflaming intergroup hostility and occasionally leading to riots (Prasad et al., 2024).[4] Electoral politics, too, has caused social rifts.[5] North 24 Parganas is adjacent to Bangladesh, a Muslim-majority country. Perceptions that Bangladeshis illegally cross the border in search of work have bolstered nativist sentiments, particularly among the majority Hindu community.[6]

### 2.2 Factory setting and production task

For the intervention, we set up our own "factory in the field." Barasat is part of the broader Kolkata industrial belt, making occasional factory work common in the area. We rented a large space in central Barasat for seven months. We hired a contractor to install tall, opaque partitions in the space, subdividing it into 23 booths. Booths had windows, matting, and a small shared table to work on. We placed fans throughout the factory, and kept work spaces clean, tidy, and comfortable throughout the study.

Workers made two types of paper bags, similar to the kind used to bag up groceries in nearby stores (see Appendix Figure S1 for pictures). The bags are made of second-hand newspaper, and are held together by a papier mâché glue (flour mixed with water and some salt, which prevents mold). There are two designs. *Small bags* are manufactured individually and are made by tearing off a strip of newspaper using a ruler, making several folds, smearing glue on the seams, and then setting aside to dry. *Big bags* must be made in pairs. They consist of an entire broad sheet of newspaper. Tears using a ruler are made on each of the long ends of the newspaper. These flaps are then folded upward and glued together. Importantly, because these flaps are large and floppy, one person needs to hold the flaps up in place while the other person applies the glue.

The big bag production function encourages a high level of simultaneous dependency in the bag-making exercise—that is, constant interactions between partners—to facilitate coordination, discussion, and perhaps friend-making. Still, even for the small bags we took steps to encourage coordination and interaction. Within each booth, pairs of workers were

[4] Amit Bhardwaj, "How the ghost of the Baduria-Basirhat communal riots is polarising polls in West Bengal," *Caravan*, April 16, 2021.

[5] The Hindu nationalist Bharatiya Janata Party has made major strides in the district in recent years, increasing its vote share from 3% in the assembly elections of 2011 to 38% in the 2021 election.

[6] "In WB's North 24 Parganas, manufactured communal hatred drives politics now," *The Wire*, April 8, 2021.

given only one small table on which to work, and only one ruler, which was necessary for making the clean tears. Partners were told that they had to jointly produce two small bags for every big bag they made, and that they had to produce these in sequence, again ensuring high inter-partner dependency.[7]

During one of the experimental waves, a research assistant unobtrusively monitored a pair of workers for ten minutes in each of the two shifts for each of the six work-days. The monitored pairs were randomly chosen in each shift. Using a stopwatch, the assistant recorded the amount of time spent by partners physically coordinating on the tasks (rather than working independently). Across the 12 ten-minute observation windows, partners physically coordinated for an average of 6 minutes and 36 seconds—thus, two-thirds of their work-time. (Non-physical interaction such as chatting occurred in addition to that.) This confirms that the task is well-suited to promoting constant engagement between partners.

We paid workers a daily wage of Rs. 400 (roughly 5 USD), for four hours of work per day (a given worker either works the 9am–1pm shift or the 1pm–5pm shift),[8] and a bonus of Rs. 200 for workers who attended all nine days of work. Among those earning income as of the baseline survey, the median experimental participant reported a daily wage of Rs. 300, paid for six working hours. Thus, given both the good working conditions and high pay, our job was seen as desirable by most people we approached with an offer in Barasat.

We hired six factory managers to oversee daily production. Their duties included tracking attendance at the start of each shift, assigning workers to booths based on our daily team lists, providing raw materials like paper and glue as needed, and recording team production at the end of each shift. They also kept the factory premises clean and tidy, and maintained a level of discipline (primarily, this meant ensuring that workers did not interact across booths).

### 2.3 Recruitment, screening, and sample selection

Enumerators recruited participants for each wave via a door-to-door campaign. Consenting individuals completed a short baseline survey in their homes, at which point they were assigned a unique ID code to bring to the factory and complete a second baseline survey

[7]We consider our intervention to have been high-intensity: it involved 24 hours of close-proximity contact between partners, as well as continuous activity and cooperation between them. There also were no external interactions or distractions during the contact period. By way of a benchmark, the median number of contact hours in the corpus of pre-registered field experiments meta-analyzed by Lowe (2025) is 2.5 hours.

[8]We decided not to pay workers a piece rate to shut down potential income confounds between treatment groups. This allows us to interpret treatment effects as being due to the different types of social interactions in the different groups, rather than income effects.

to express their interest. Upon completing the second baseline survey, they were assigned to a three-day lead-in period, where they worked individually for roughly four hours per day. For these three days, each participant shared a booth with one or two ingroup participants.

Only those who completed all three days of the lead-in period were then randomized to one of the treatment groups described below. This helped us to screen out individuals who were less likely to follow through with the intervention, reducing attrition. Of the 1,479 men who completed the first baseline, 85% completed the second baseline, and of those, 77% completed the lead-in period. The multistage recruitment process for the final experiment also allows us to explicitly examine selection at each stage. We find that participants who make it to the final sample are somewhat more educated and less likely to be employed. However, there is no evidence of selection based on prejudice or political attitudes, which ex ante might have been a concern (Appendix Table S1).

Factory managers trained the workers twice—on the initial day of the lead-in period on making small bags, and then on the day workers began working in pairs, focusing on the production of the big bags that required both workers to collaborate.[9]

### 2.4 Randomization

We assigned the 964 participants who completed the three-day lead-in period to one of four intervention arms, stratifying by intervention wave (one to ten) and religion (Hindu versus Muslim, visualized in Figure 1):

1. *Outgroup, broad.* Participants were assigned to work together in their booth with a different outgroup partner for each day of the six-day intervention period. The six outgroup partners were randomly chosen (and randomly ordered) from among the outgroup members assigned to *Outgroup, broad* in the same wave and shift. Participants were informed at the start of the intervention period that their partners would change daily. However, they were not informed about their partners' identities, nor were they told that their partners would be outgroup members.
2. *Outgroup, deep.* Participants were paired with a single outgroup member, randomly chosen from those assigned to *Outgroup, deep* in the same wave and shift, for the entire six-day intervention period. They were informed on the very first day that they

[9] Since the factory managers were not professional bag-makers, we brought in an expert to train them first. The expert instructed the managers on bag-making techniques and taught them how to rate the quality of the bags produced.

would be paired with the same person throughout, but again no mention was made of their partners' identities.

3. *Ingroup, deep.* This group was similar to the *Outgroup, deep* treatment except for the fact that participants were always with an ingroup member and this treatment arm consisted solely of Hindu workers—all pairs were therefore Hindu-only teams. This treatment provides a benchmark for assessing whether day-to-day interactions with an outgroup partner can reasonably be characterized as deep and positive. [10]
4. *Pure control.* After the three-day lead-in period, these participants were randomized to serve as the pure control group—they did not participate in factory work during the following six days, but still responded to the subsequent endline survey.[11]

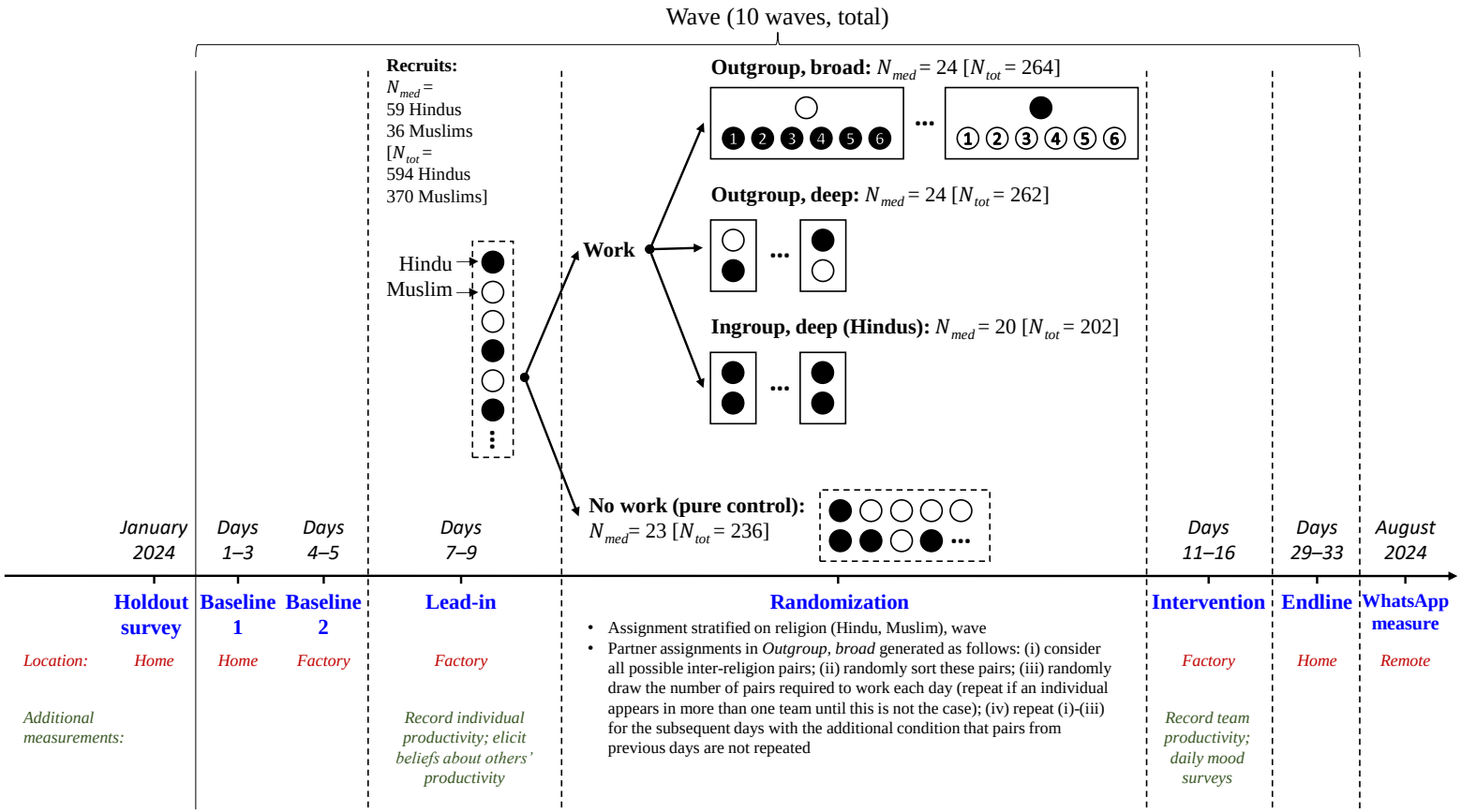
Our core design innovation is to fix the total hours of intergroup contact ($N = 24$), while varying the number of outgroup members met ($n \in \{1, 6\}$). With unlimited sample and budget, we would vary $N$ and $n$, which could yield additional insights—for example, some information about a partner may only be revealed after years of friendship. In practice, we selected parameters that are within a policy-maker's feasible set of contact interventions, although at the upper end of treatment intensity—at the 84th percentile of pre-registered contact interventions for depth (hours of contact per outgroup member met), and at the 92nd percentile for breadth (number of outgroup members met), as shown in Figure S2.[12] Otherwise, our parameter selection intended to navigate tradeoffs and constraints in reasonable ways—for example, budget constraints help cap $N$, given the financial costs of keeping participants fully employed for $N$ hours, and power considerations help determine our $n \in \{1, 6\}$, which maximizes variation subject to the logistical constraint of no within-day partner changes.

[10]Note that this treatment group is the closest to the form of the three-day lead-in period. There are two differences: in the lead-in period some participants had two booth partners as opposed to one, and lead-in period production was all done individually (i.e., it involved no production of big bags).

[11]We did not include an *Ingroup, broad* condition due to power concerns: adding the minimum required number of individuals (for six days of contact) to this condition per wave (while holding the overall sample size fixed) would have required reallocating 25% of Hindus from outgroup treatments, which we viewed as a significant cost given our sample constraints. The lack of an *Ingroup, broad* condition has minimal drawbacks given our core questions of interest. In particular, we do not need an *Ingroup, broad* condition to cleanly identify the effects of broad versus deep contact, holding total contact time fixed (by comparing *Outgroup, broad* with *Outgroup, deep*); nor do we need an *Ingroup, broad* condition to identify the effects of both types of contact relative to the *Ingroup, deep* benchmark. The one drawback relates to pinning down mechanisms: strictly speaking, we cannot rule out that effects of *Outgroup, broad* relative to *Outgroup, deep* are partially due to the effects of meeting more people *in general*, rather than specifically the effects of meeting more outgroup members. However, the latter is the more plausible channel, given the outcomes we find effects on (e.g. beliefs about outgroup strangers) and those that we do not find effects on (e.g. beliefs about ingroup strangers).

[12]The Figure compares our treatments to the breadth and depth of contact in the pre-registered intergroup contact experiments reviewed in Lowe (2025).

Figure 1: Timeline and randomization.




*Notes:* This figure shows the timeline of the experiment and the randomization scheme, visualized for one experimental wave. $N_{med}$ denotes the median sample size for a given group in each wave. $N_{tot}$ denotes the total sample size for a given group across all waves of the experiment. Sample sizes only include the 964 subjects who completed all three days of lead-in work, and thus were randomized to an experimental condition. Totals include subjects who did not complete the endline survey. Rectangles with solid lines represent partner pairings—a different pairing for each intervention day in *Outgroup, broad*, and the same pairing for all intervention days in *Outgroup, deep* and *Ingroup, deep*.

To assign partners in the *Outgroup, broad* condition, we used the following algorithm:

1. Consider all possible inter-religious pairs in the *Outgroup, broad* arm.
2. Randomly sort these pairs.
3. Randomly draw the number of pairs required to work each day; repeat if an individual appears in more than one pair until this is not the case.
4. Follow steps 1 to 3 above for the next days with the additional condition that pairs from previous days are not repeated.

After randomly assigning workers to partners, we randomly assigned worker-pairs to booths for each day of the intervention. This means that regardless of whether the partner is fixed or rotating, a given worker would generally have to move to a different booth each day. Otherwise, if a worker had a missing partner on any particular day, they were asked to make small bags only.

Between two to three weeks after a given wave's six-day work period was completed (median gap of 18 days), enumerators visited the homes of the participants to administer the endline survey. One month after the final endline survey, we sent a WhatsApp message to each participant, recording whether they replied—we explain the purpose of this measure in detail in the next section.

Note, in the analysis that follows, our primary focus is on the comparison of *Outgroup, deep* versus *Outgroup, broad*, which cleanly isolates differences in the depth versus breadth of outgroup exposure while holding total contact time fixed. Where relevant to specific theoretical claims, we also draw on auxiliary comparisons with the pure control and *Ingroup, deep* conditions, helping to clarify mechanisms and benchmark effects.

# 3 Data and empirical strategy

## 3.1 Outcomes

**Daily mood surveys** At the end of each of the six main intervention period workdays, workers completed mood surveys. Workers reported how they felt that day on a five-point scale [Very unhappy / Somewhat unhappy / Neither happy nor unhappy / Somewhat happy / Very happy]. They reported how attached they felt to their partners using the *Inclusion of Other in the Self* scale introduced by Swann et al. (2009) (a five-point scale ranging from a pictorial representation of the two partners as separate circles, to one of fully overlapping

circles, see Appendix Figure S3). We asked "How close of a friend do you consider your partner to be?" on a ten-point scale (Not at all close to Very close). We also asked participants to rate their own performance at work that day as well as their partner's performance on a ten-point scale (Very bad to Very good). On the final day of work, we asked participants more detailed questions about the topics they discussed with their partners during work.

**Productivity** At the end of every shift, factory managers went to each booth to count the number of big and small bags produced by each pair of workers. They also rated the quality of the bags on a scale of 0 to 10—one score for the big bags produced in that booth in that shift, and two scores for the small bags produced, with one score for those produced by each worker in the booth.

On the third and final day of the lead-in period (before randomization into treatments), we presented workers with five pairs of names (four mixed-religion pairs[13] and one same-religion pair) from participants in the previous wave and asked them to predict which worker within each pair produced more bags per hour during the lead-in period. One of their predictions was randomly chosen for a correct-guess reward of Rs. 50. We also asked them to estimate the difference in productivity for each pair, earning an additional Rs. 10 if their estimate was within two bags of the actual difference. These data allow us to measure baseline misperceptions about outgroup members' productivity.

**Social and economic ties** At endline we asked participants assigned to the work treatments whether they had had any contact over the past week with any of their intervention-period work partners. Among those that answered yes, participants selected which types of contact they had had: whether they had seen each other in person, taken food or drink together, spent time with each other's families, or called or sent messages via phone. Broad participants answered this question separately for their partners from days one, four, and six. For a more objective measure of social ties with partners, we asked participants to tell us the last four digits of their assigned work partner's mobile phone number, if they had it, and we match responses to the partner's phone numbers provided at baseline (88% of reported numbers match a partner's baseline-provided number). Again, we asked broad participants only about their partners from days one, four, and six.

[13]The four mixed-religion pairs featured individuals from the 25th, 50th, 75th, and 100th productivity percentiles within their groups for each wave. This meant that a Hindu worker from the 25th percentile was paired with a Muslim worker from the 25th percentile, and the same pairing method was applied for the other percentiles.

To measure economic ties, we asked participants assigned to the work treatments whether (i) any of their work partners had helped them to find other paid work, and (ii) whether they had borrowed from or lent money to any of their work partners. For both questions, when a broad participant answered yes, we asked them to answer the same question separately for each of their partners from days one, four, and six.

**Outgroup misperceptions** Before starting the experiment, we carried out a survey of 83 Muslims and 125 Hindus in the same area, to be used to benchmark the incentivized beliefs of our experimental participants. The sampling protocol was the same as the one we later used to recruit the experimental sample. After visiting people at their homes, we invited them to answer questions and play some experimental games. We recorded some basic information about the individuals, including their first names (see Appendix Table S13 for summary statistics). Depending on how they responded to the questions and games, subjects were able to earn a small amount of money. We refer to this pool of subjects as the "holdout sample" since they did not participate in the study in any other way.

Later, in the endline survey, we asked experimental subjects to guess the preferences and behaviors of one Hindu and one Muslim individual randomly drawn from the holdout sample, in random order.[14] We carefully described the games to subjects. To avoid priming subjects about the study's focus on intergroup relations, we did not give the religion of the holdout subject being guessed about, only their first name. To ensure that the individuals being guessed about were recognizable as Hindus or Muslims, we only asked about holdout subjects whose first names unambiguously indicated their religion (68 Muslims and 97 Hindus).[15]

We collected nine measures of misperceptions (Table 2). We chose misperceptions that are common locally—based on our prior knowledge of the area and informal discussions. We make extensive use of behavioral games to help obfuscate the underlying concept being targeted (for example, to measure hygiene preferences we ask holdout subjects if they would like to use hand sanitizer after handling some dirty coins, and record whether they take some).

In reality Hindus and Muslims behave similarly (Figure 2). The only significant differences we see are that Muslims cheat less than Hindus (reporting 5.72 versus 6.46

[14]The one exception is for guesses about productivity, where we had endline subjects guess about the productivity of a worker from a different experimental wave.

[15]For the guesses about non-binary measures (e.g. productivity), to make guessing easier, we asked questions in two steps. First, we asked the subject to guess whether (i) the Hindu and Muslim behaved the same, (ii) the Hindu scored higher (e.g. higher productivity), or (iii) the Muslim scored higher. In the second step, we asked the subject to guess the number for each person separately, with the survey flow delivering an error message if these second-step answers were inconsistent with the first-step answer.

Table 2: Dimensions of outgroup misperceptions and associated measurements.

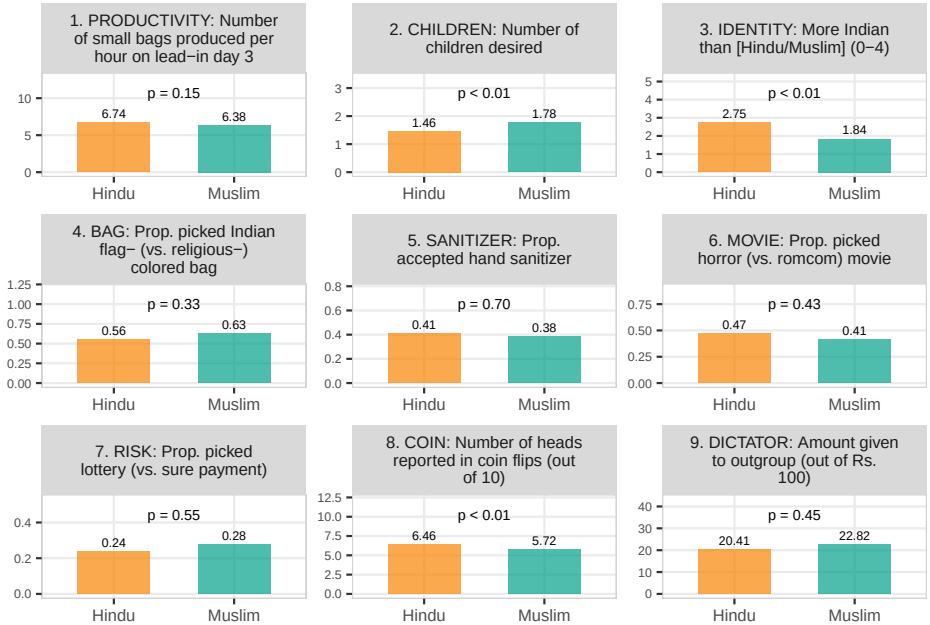
| Concept | Measure of "truth" | Measure to elicit beliefs |
|---|---|---|
| Productivity | Number of small bags an outgroup member produces per hour during the lead-in period | How many small bags per hour do you guess that [NAME] made during his third day at the factory? (Correct +/- 3 bags = Rs. 5) |
| Fertility preferences | How many children would you like to have in total? | How many children do you think that [NAME] reported that he would like to have in total? (Correct +/- one child = Rs. 5) |
| Identity attachment to religious group versus nation | Suppose that you had to choose between being an Indian and being a [Hindu / Muslim]. Which of these two identities do you feel most strongly attached to? | Do you think that [NAME] said he felt himself to be more Indian or more a member of his religious group? (Correct answer = Rs. 5) |
| | Does respondent choose a [green / saffron] or an Indian flag-color bag as a gift? [Hindus choose between saffron or flag-color, Muslims choose between green or flag-color] | Do you think [NAME] chose the [green / saffron] bag or the Indian flag-color? (Correct answer = Rs. 5) |
| Hygiene | Does respondent take hand sanitizer when offered after handling dirty coins? | Do you think that [NAME] used the hand sanitizer? (Correct answer = Rs. 5) |
| Courage | Does respondent choose horror or romantic comedy film as a gift? | Do you think that [NAME] chose the horror or the romantic comedy film? (Correct answer = Rs. 5) |
| Risk preferences | Does respondent choose to enter a lottery or get a sure payment? | Do you think that [NAME] chose the risky lottery draw or the sure payment? (Correct answer = Rs. 5) |
| Honesty | Respondent flips 10 coins in private; gets paid for number of heads they self-report (incentive to cheat) | How many heads from 0 to 10 do you think that [NAME] reported? (Correct +/- one flip = Rs. 5) |
| Altruism | Giving to outgroup stranger in a dictator game | How much of the Rs. 100 do you think [NAME] gave away to a stranger belonging to a different religious community, keeping the rest for himself? (Correct +/- Rs. 10 = Rs. 5) |

heads), prefer more children (1.78 versus 1.45), and identify more with their religion, as opposed to national identity (although Muslims are somewhat more likely to pick an Indian-flag colored bag as a gift than a religion-colored bag). These differences fit with the idea that stereotypical beliefs, like the belief that Muslims have larger families than Hindus, can contain a kernel of truth (Bordalo et al. 2016).[16]

Our core endline-derived measure of outgroup belief accuracy is an index of nine binary components, each assessing how close endline subjects' guesses were to the response of the average outgroup member in the holdout sample (e.g., if the experimental subject was guessing about a holdout subject with a Muslim name, did they successfully predict the behavior of the average Muslim subject in the holdout sample). We dichotomize the guesses as right or wrong using the cutoffs used for the incentive payments shown in the third column

[16]Muslims in India are often stereotyped in political and media discourse as being less patriotic or as having disproportionately high fertility. For example, see illustrative coverage in Al Jazeera's *"From Hamas to Halal: How anti-Muslim hate speech is spreading in India"* (Feb 26, 2024) and Le Monde's *"In India, Islamophobia gains ground in election campaign"* (Apr 24, 2024).

Figure 2: Preferences and behavior of holdout sample, by religion.




*Notes:* This figure presents the average responses, by religion, to survey questions and behavioral games carried out by participants in the holdout survey (panels 2–9). Panel 1 presents the average number of bags made per hour on the final lead-in day. Only participants whose names were used for guessing in the endline are included in the sample.

of Table 2.[17] We then collapse these nine binary variables into an index by standardizing each and then taking the simple average (following Kling et al. 2007).

For some analyses, we distinguish between guesses about more versus less visible attributes. We classify less-visible attributes as those that are abstract in nature and relatively hard to discern in the context of everyday factory work: risk preferences, courage, honesty, and altruism. We classify more-visible attributes as those that can be readily inferred at work: productivity (since this is directly observed), fertility preferences (since family is a ubiquitous conversation topic in Bengal, and the second-most common topic of conversation at work, as we show below in Table 4), our two measures of religious identity (since whether or not partners pray and/or don religious jewelry or clothing such as skull caps is easy to observe), and hygiene (since booths are small, and partners' hygiene can be seen or smelled).

[17]The one deviation from this cutoff rule is for the number of bags produced. In the endline survey we asked subjects to guess how many bags per hour an outgroup member made on the final lead-in day, and said that they would receive a payment if they were correct within three bags. The mean number of bags produced was 6.58, meaning that the vast majority of subjects guessed within three bags, giving us little variation. Thus, for the main measure, we say that a guess is correct only if it is within one bag of the truth. In Appendix Table S9, we show that our results are robust to using continuous measures for the four components that can be measured as continuous absolute differences between the guess and the truth benchmark (coin toss, desired fertility, productivity, and dictator giving).

While reasonable, we caveat that this division of attributes into high versus low visibility was not pre-registered, and thus results split by visibility should be considered more exploratory.

Finally, for some attributes, we also elicit predictions about a worker's partner, permitting a test of the model's prediction that deep contact leads to more accurate beliefs about the partner than broad contact. This belief data is more limited than our measures of beliefs about strangers. We discuss details in Subsection 5.4.

**Generalized intergroup behaviors** We measure generalized behavior toward the outgroup using three endline survey measures and one post-endline "audit" measure. First, to explore effects on close friendships, we asked participants to report the last names of their five closest friends. Our main outcome is the number of close outgroup friends, coded using the religion signalled by the last names.

Second, to measure altruism, all participants were randomly matched with an outgroup member from the control group and asked to split Rs. 100 between themselves and their partner in a dictator game. The religion of their partner was made salient simply by their first name.

Third, to measure willingness to interact with outgroup strangers, we informed participants that we would invite 20 people from Barasat for an additional day of paid work in the future. Once again, we randomly matched them with a named outgroup member from the control group and asked if they were willing to work for Rs. 700, then Rs. 600, Rs. 500, and so on down to Rs. 0. As soon as the respondent refused an offer, we stopped asking further. We define the reservation wage as the midpoint between the two wage offers that cause a participant to change their answer from yes to no. If they always say yes or always say no, we have upper and lower bounds for the reservation wage (0 and 700). We then use tobit estimation with censoring to estimate effects on the reservation wage. We interpret a negative effect on the reservation wage as a positive effect on willingness to work with outgroup strangers.

Finally, for an obfuscated measure of civility toward the outgroup, we sent each participant a WhatsApp message from an outgroup stranger (an unknown employee of the factory), as signalled by their name (for details see Appendix C.7), roughly one month after the final endline survey. We estimate effects on whether or not the participant replies to the message.

### 3.2 Summary statistics, balance, and attrition

We present summary statistics for the experimental sample in Appendix Table S2. The sample is young (average age of 22) and 87% are unmarried. 65 percent of subjects have completed schooling up to 12th grade, though few (12%) have completed university degrees. 62% of subjects are Hindus and 38% are Muslims. Only 23% of subjects are currently employed, which is to be expected since it was clearly stated that participating in the study required almost two weeks' worth of time during workdays (Monday to Saturday). We find significant demand for social interaction: 67% reported that they would be very interested in attending a social event to meet new people of their age and gender. Subjects report few depressive symptoms (a mean of 2.4 on a 24-point PHQ-8 depression score).

Our sample is well-balanced on pre-treatment covariates: six of 147 tests are significant at the 5% level, consistent with chance (Appendix Table S3). Ninety-five percent of subjects completed the endline survey, and attrition is uncorrelated with treatment assignment and religious group (columns 1 and 2, Appendix Table S4). Work attendance averaged 96%, and was also similar by religious group (column 4). Attendance was marginally significantly higher in *Outgroup, broad* than in *Outgroup, deep* ($p = 0.06$, column 3), but the gap of two percentage points off a base of 94% is economically small.

### 3.3 Baseline bias and exposure

Participants exhibit ingroup bias at baseline (Appendix Figure S4).[18] When ranking the trustworthiness of five groups (1 = most trustworthy), Muslims rate Muslims at 1.1 on average and Hindus at 2.5—a gap of 1.4. Hindus, by contrast, rate Hindus at 1.1 and Muslims at 3.4—a gap of 2.3. Thus, the majority distrusts the minority more than vice versa. India's Hindu-nationalist Prime Minister, Narendra Modi, is viewed favorably overall (mean feeling-thermometer score = 67/100), but attitudes diverge sharply: Hindus average 85, whereas Muslims average 39.

Turning to misperceptions, we only measured baseline beliefs about productivity.[19] Immediately after the lead-in period and before treatment began, we elicited incentivized predictions of the productivity of randomly selected Hindu and Muslim participants from the previous wave. We provided each participant with four pairs of mixed names (one Hindu

[18] Note that we did not want to prime participants into believing that the study centers on intergroup relations, so we kept baseline measurements of bias and prejudice to a minimum.

[19] We chose to measure baseline misconceptions using productivity (rather than more politically sensitive outcomes like fertility or national identity) in order to minimize priming.

and one Muslim in each) plus one same-religion pair. We told respondents each pair's total hourly output (the sum of individual outputs) and asked them to predict their individual productivity for a reward. Hindus have considerably more ingroup-biased beliefs about productivity than Muslims (Appendix Figure S5).

Finally, in Section 5.3 we show that at endline, control group Hindus exhibit larger misperceptions about Muslims than the reverse (using the combined index of nine belief dimensions explained above). Overall, our sample displays meaningful bias and prejudice, with the asymmetry that the majority shows more ingroup bias than the minority. The latter is consistent with the fact that the majority has much less baseline exposure (both at work and in their home neighborhood) to the minority than the other way around, as we show in Appendix Figure S7.

### 3.4 Pre-registration and demand effects

We pre-registered our experiment in the AEA RCT Registry (#13126) during the first wave of the experiment. At the time of registration, 14 subjects had completed the endline survey, but no data had been accessed by any of the research team. We did not pre-register the distinction we draw between visible and non-visible attributes—this distinction was introduced ex post to provide a lens for interpreting heterogeneous effects across belief outcomes. We detail other minor deviations from the pre-registration in Appendix C.1.

Given the array of outcomes we measure, and resultant array of hypothesis tests, Appendix Table S14 reports sharpened false discovery rate (FDR) $q$-values (Benjamini et al., 2006; Anderson, 2008) for our 35 main broad versus deep comparisons—those summarized in Table 1. Under the sharpened FDR procedure, rejecting hypotheses with $q < 0.1$ controls the expected proportion of false discoveries among rejections at 10%. We report multiple-testing adjustments both within outcome domains and across all 35 comparisons jointly. Every broad-versus-deep difference in Table S14 with a statistically significant unadjusted $p$-value has $q < 0.1$ within its domain. Treating all 35 comparisons as one family, three outcomes become insignificant ($q > 0.1$): happiness at work ($q = 0.13$), happiness at work on day one ($q = 0.18$), and whether the work partner is considered a close friend on day one ($q = 0.12$). Accordingly, our headline findings (i.e., those emphasized in our abstract) do not involve these three outcomes.

We took several steps to mask the study's focus on intergroup relations. We avoided any mention of religion in both factory operations and the endline survey, using religiously distinctive names instead. In the final endline question, only 11% of respondents guessed

that the study was about Hindu-Muslim relations (Figure S8), and guesses do not vary by treatment status (Table S5). The vast majority of participants thought that the study was about jobs, employment, or other unrelated issues. Consequently, the pattern of results is unlikely to reflect experimenter demand effects.

### 3.5 Empirical specification

We primarily use the following specification:

$$y_i = \psi_s + \beta_{ID} \textit{Ingroup, deep}_i + \beta_{OD} \textit{Outgroup, deep}_i + \beta_{OB} \textit{Outgroup, broad}_i + \varepsilon_i \quad (1)$$

where $y_i$ denotes the outcome recorded for individual $i$, and $\psi_s$ are 20 randomization strata (religion $\times$ wave) fixed effects. *Ingroup, deep*$_i$, *Outgroup, deep*$_i$, and *Outgroup, broad*$_i$ are binary indicators for individuals' treatment status, with the pure control group as the omitted category. $\beta_{ID}$ is identified only from Hindu participants, since no Muslims are assigned to the *Ingroup, deep* condition. For simplicity, some figures compare only individuals in the *Outgroup, deep* and *Outgroup, broad* conditions when the key comparison is between types of outgroup exposure. Comparisons with the pure control and *Ingroup, deep* conditions, when relevant, are shown in the Appendix for these cases. We report robust standard errors, as pre-specified.[20]

For analyses employing daily mood surveys, we estimate this specification:

$$y_{i,t} = \mu_s + \gamma_{OD} \textit{Outgroup, deep}_i + \gamma_{OB} \textit{Outgroup, broad}_i + u_{i,t} \quad (2)$$

Outcomes are measured daily, with $t \in \{1,2,3,4,5,6\}$, and *Ingroup, deep* serves as the omitted category, since pure control group respondents do not take the mood surveys. The strata ($\mu_s$) are religion $\times$ wave $\times$ day $\times$ shift fixed effects (recall that the religion $\times$ wave denotes the randomization strata). Standard errors are clustered by individual. We drop observations for individuals whose partner was absent on that day.

Finally, for analyses with team productivity as the outcome, we use the following specification:

$$y_{k,t} = \phi_s + \alpha_{OD} \textit{Outgroup, deep}_k + \alpha_{OB} \textit{Outgroup, broad}_k + \psi_{k,t} \quad (3)$$

[20]We note that the $p$-values for our core results are similar when clustering standard errors at the partner-network-level (pair for deep partners, 12-person rotation batch for broad partners, Appendix Table S14). Across 35 tests, the median p-value change from clustering is only an increase of 1.1 percentage points. No test that is significant at the 10% level loses significance after clustering.

Team productivity, like mood, is measured daily, but we have one observation per team $k$ (instead of per individual) each day. The strata ($\phi_s$) are wave $\times$ day $\times$ shift fixed effects. Standard errors are clustered at the team level.

When available, we also include baseline controls in our specifications, such as each partner's individual productivity during the lead-in period.

# 4 Conceptual Framework

We propose a simple model of how contact shapes belief updating, distinguishing between learning about an individual partner and generalizing from those experiences to the outgroup as a whole. Readers primarily interested in the empirical findings may wish to skip ahead to Section 5.

The framework proceeds in two steps. Section 4.1 presents the core sampling problem. Taking the information generated by an interaction as given, it derives the trade-off between breadth and depth: repeated interactions with the same person can improve learning about that individual, while interactions with more distinct individuals provide more independent observations and therefore facilitate learning about the wider group. The model also predicts diminishing returns to additional contact, so individuals with greater prior exposure to the outgroup should update less in response to further contact. Our experiment was designed around these core predictions.

Section 4.2 extends the framework to organize heterogeneity observed ex post across outcomes by allowing the information generated by an interaction to depend on relational investment. Some characteristics are readily *visible* during ordinary interaction, whereas others may require greater relational investment before they can be learned. For attributes that require such investment, depth can affect not only the precision of information but also whether particular information is acquired at all, because investment is more attractive when future interaction is anticipated.

## 4.1 Visible attributes and sampling

While the attribute averaged across all outgroup members is $x^* \in \mathbb{R}$, the outgroup population is heterogeneous and the true attribute of any outgroup individual $i$ is $x^* + \eta_i$ where $\eta_i$ is the $i$-specific mean-zero deviation from the group mean. The DM starts with an inaccurate estimate $x_0^{st} \neq x^*$ about $x^*$. She believes that $x_0^{st}$ is unbiased, i.e., $x_0^{st} = x^* + e^{st}$, where $e^{st}$ is mean-zero

noise with variance $\sigma^2_{st}$.[21] Thus, the DM knows $x^{st}_0$ but not $x^*$.[22] We assume that the DM interacts with $n \geq 1$ outgroup members (contact *breadth*), and meets each of them $D \geq 1$ times (contact *depth*). The DM thus has $N = nD$ total interactions with outgroup members.

**Noisy observations** If the attribute is *visible* during interactions, then the DM can observe and use it to gradually learn about the attribute. For example, while working together in a factory, a work partner's productivity or propensity to wear religious symbols is physically visible and hence can be learned, but a partner's altruistic tendencies are not physically visible and hence cannot be immediately learned. At any interaction with outgroup member $i \leq n$ on occasion $d \leq D$, the DM only observes $x_{id} = x^* + \eta_i + \varepsilon_{id}$ where $\varepsilon_{id}$ is an individual-occasion specific noise term. To study how the DM updates about a generic outgroup member that the DM never encounters, we will write $x^* + \eta_m$ as the true attribute of such an individual $m \neq 1, 2, ..n$.

**DM's model of the world** We will use $\varepsilon$ and $\eta$ as a shorthand for the collection of all random variables $(\varepsilon_{id})_{i,d}$ and $(\eta_i)_i$ respectively. Let $E$ denote the expectation with respect to the joint distribution of all $e^{st}, \varepsilon, \eta$ variables, for *all* individuals including $m$. To make the notation simpler, for any random variables $X, Y, Z$, we will use $E(X|Y,Z)$ to denote the expectation of $X$ conditional on $Y, Z$. The DM makes the following assumptions about the world:

- $E(e^{st}|\varepsilon, \eta) = 0$, $E((e^{st})^2|\varepsilon, \eta) = \sigma^2_{st}$.
- $E(\eta_i|e^{st}, \varepsilon) = 0$, $E(\eta_i^2|e^{st}, \varepsilon) = \sigma^2_{het}$, and for $i \neq j$, $E(\eta_i \eta_j|e^{st}, \varepsilon) = 0$.
- $E(\varepsilon_{id}|e^{st}, \eta) = 0$, $E(\varepsilon^2_{id}|e^{st}, \eta) = \sigma^2_{inf}$, and unless $i_1 = i_2$ and $d_1 = d_2$, $E(\varepsilon_{i_1 d_1} \varepsilon_{i_2 d_2}|e^{st}, \eta) = 0$.

These assumptions imply that all error components have unconditional mean zero and that distinct error components are pairwise uncorrelated. $\sigma^2_{inf}$ measures how noisy the observations are and $\sigma^2_{het}$ measures how heterogeneous the outgroup attribute is. The DM uses these assumptions to form a guess $\widehat{x_{n,D}}$ of the attribute for an outgroup individual $i$.

Next, we define a notion of accuracy that benchmarks the DM's guess $\widehat{x_{n,D}}$ against the true value, $x^* + \eta_i$. Both bias (uni-directional mistakes stemming from consistently skewing towards $x^{st}_0$) and observational noise can cause the guess to deviate from the truth, thereby reducing accuracy.

[21] One could also assume that in reality, the DM has a biased estimate, i.e., $x^{st}_0 = x^* + c + e^{st}$ where $c$ is a constant bias that the DM is unaware of. This would not affect any of the stated results that follow.

[22] While the DM treats $x^{st}_0$ as a random unbiased draw, we, as analysts, treat it as a constant.

**Definition 1.** *The inaccuracy $\Delta$ when guessing the attribute for person i is defined as the expected squared distance between $\widehat{x_{n,D}}$ and the truth, that is $\Delta(\widehat{x_{n,D}}) = E(\widehat{x_{n,D}} - x^* - \eta_i)^2$. The accuracy $\mathscr{A}$ is just the negative of inaccuracy: $\mathscr{A}(\widehat{x_{n,D}}) = -\Delta(\widehat{x_{n,D}})$.*

**DM's Guess for any** $n, D$ To update his guess optimally about a generic outgroup member $m$, the DM assigns weights $w^{st}, \{w_{id}\}_{i,d}$ to each corresponding piece of information, $x_0^{st}, \{x_{id}\}_{i,d}$, to maximize the accuracy of the weighted "unbiased"[23] linear estimator

$$\widehat{x_{n,D}}(w^{st}, (w_{id})_{i,d}) = w^{st} \times x_0^{st} + \sum_{i \leq n} \sum_{d \leq D} w_{id} \times x_{id}, \text{where, } w^{st} + \sum_{i \leq n} \sum_{d \leq D} w_{id} = 1$$

While the linear estimator might seem restrictive, all our results generalize in a fully Bayesian updating framework, as we discuss at the end of this section. The DM optimally updates his guess about a generic outgroup member $m$ as

$$\begin{aligned}
\widehat{x_{n,D}}^{gen} &= \arg\max_{\widehat{x_{n,D}}} -E\left(\widehat{x_{n,D}}(w^{st}, (w_{id})_{i,d}) - x^* - \eta_m\right)^2 \\
&= \arg\max_{\widehat{x_{n,D}}} -E\left(w^{st}(x_0^{st} - x^*) + \sum_{i \leq n} \sum_{d \leq D} w_{id}(x_{id} - x^*) - \eta_m\right)^2 \\
&= \arg\max_{\widehat{x_{n,D}}} -E\left(w^{st}(x_0^{st} - x^*) + \sum_{i \leq n} \sum_{d \leq D} w_{id}(x_{id} - x^*)\right)^2 \\
&= \arg\max_{\widehat{x_{n,D}}} -\Bigg((w^{st})^2 E(e^{st})^2 + \sum_{i \leq n} \sum_{d \leq D} (w_{id})^2 E(\eta_i^2 + \varepsilon_{id}^2) \\
&\qquad + 2\sum_{i \leq n} \sum_{1 \leq d' < d \leq D} w_{id} w_{id'} E(\eta_i^2)\Bigg) \qquad (4) \\
&= \frac{1}{\frac{1}{\sigma_{st}^2} + \frac{n}{\sigma_{inf}^2/D + \sigma_{het}^2}} \left(\frac{1}{\sigma_{st}^2} \times x_0^{st} + \sum_{i \leq n} \frac{1}{\sigma_{inf}^2/D + \sigma_{het}^2} \times \frac{\sum_{d \leq D} x_{id}}{D}\right) \qquad (5)
\end{aligned}$$

Note that this expression is calculated from the DM's perspective, and hence uses the DM's assumption $x_0^{st} - x^* = e^{st}$. Appendix C.2 contains the complete derivation of equation (5). To provide an overview of the proof, we present a rough outline for the steps leading to the final expression in (5). Specifically, because $\eta_m$ is uncorrelated with every other piece of information, it becomes irrelevant to the optimal weighting rule. Next, we substitute $w^{st} = 1 - \sum_{i,d} w_{id}$ into the objective and differentiate with respect to the $nD$ observation weights. Invariance to permutations of individuals and to permutations of occasions within

[23] Given the DM believes that $Ex_0^{st} = x^*$, unbiasedness from the DM's perspective requires $w^{st} + \sum_{i,d} w_{id} = 1$.

each individual then implies equal observation weights.

Overall, as shown in equation (5), maximizing accuracy $\mathscr{A}$ requires weighting each *uncorrelated* piece of information, be it the initial measure $x_0^{st}$, or each of the $n$ sample averages from $n$ outgroup members, $\frac{\sum_{d\leq D} x_{id}}{D}$, by the inverse of its corresponding variance $\sigma_{st}^2$ and $\sigma_{inf}^2/D+\sigma_{het}^2$ respectively. The weights are then normalized (see preceding fraction that has 1 in the numerator) so that they sum to 1. With such inverse variance weights, more precise information gets more importance.[24] For example, when the prior is considered to be noisier (higher $\sigma_{st}$), or when the observations are more precise (lower $\sigma_{inf}$), or when the population is less heterogeneous (lower $\sigma_{het}$) and hence every encountered individual is more informative about individual $m$, then the prior estimate $x_0^{st}$ loses importance.

Similarly, after interacting with $n$ individuals, the DM forms the following optimal guess about *any one* of the $n$ individuals he has interacted with (say individual $I$):

$$\begin{aligned}\widehat{x_{n,D}}^I &= \arg\max_{\widehat{x_{n,D}}} -E\left(\widehat{x_{n,D}}(w^{st},(w_{id})_{i,d})-(x^*+\eta_I)\right)^2\\ &= \arg\max_{\widehat{x_{n,D}}} -E\left(w^{st}(e^{st}-\eta_I)+\sum_{i\neq I}\sum_{d\leq D} w_{id}(\eta_i+\varepsilon_{id}-\eta_I)+\sum_{d\leq D} w_{Id}\varepsilon_{Id}\right)^2\end{aligned}$$

For $i\neq I, d\leq D$, this solves as:

$$\begin{gathered} w_{id}=\frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)},\quad w_{Id}=\frac{\sigma_{st}^2\sigma_{inf}^2+nD\sigma_{st}^2\sigma_{het}^2+\sigma_{inf}^2\sigma_{het}^2+D\sigma_{het}^4}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)},\\ w^{st}=\frac{\sigma_{inf}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\end{gathered} \tag{6}$$

Appendix C.3 contains the complete derivation. Even though the DM has encountered $n$ individuals in both cases, the guess $\widehat{x_{n,D}}^I$ about an encountered person differs from the guess about the general outgroup $\widehat{x_{n,D}}^{gen}$. In particular, the observations $x_{Id}$ are informationally favored over all $x_{id}$ with $w_{Id}>w_{id}$, as only the former informs about $\eta_I$. Relatedly, the error/noise in $x_0^{st}$ becomes correlated to that in the sample average obtained from any non-$I$ outgroup member, as both fail to account for $\eta_I$.[25] The mathematical steps in deriving (6) are identical to the last case: calculate accuracy, calculate its derivatives with respect to the weights, and then solve the system of derived equations.

[24] The same class of estimators is also used in meta-studies, where larger studies with more precise effect size estimates are given more weight and smaller, noisier studies are given less weight.

[25] For comparison, when updating about the stranger $m$, all information sources were equally (un)-informative about $\eta_m$.

**Results** The following results stated formally below establish our main hypotheses: for visible attributes, as long as the initial perception $x_0^{st}$ is inaccurate enough, both broader contact (increasing $n$) and deeper contact (increasing $D$) enhance the accuracy about the outgroup at large, but with *diminishing returns*. Thus, on the margin, the accuracy-increasing effect of increased contact (higher $n$ or $D$) is smaller for individuals who have already encountered a higher number of outgroup members or encountered them for longer, respectively. In our setting, the diminishing returns with respect to $n$ means that effects should be lower for Muslims, the minority group, than Hindus.

Next, when one needs to choose between broader or deeper contact, the former is preferable to increase the accuracy about a random outgroup stranger (subject to $x_0^{st}$ being inaccurate enough), but the latter is preferable to increase the accuracy about an encountered outgroup member (subject to the outgroup heterogeneity $\sigma_{het}$ being large enough). Thus, there is a *sampling trade-off* between how much the DM can update about each encountered person, and how much they can generalize that to the wider group. This tradeoff result will play a crucial role in Section 4.2.

While the DM interprets $x_0^{st}$ as a draw with mean-zero noise and variance $\sigma_{st}^2$, we as the analyst treat $x_0^{st}$ as a constant. Therefore, the DM's perceived $\sigma_{st}$ influences our results solely through its role in determining the DM's updating rules, (5) and (6). Let $C = (n, D, \sigma_{st}, \sigma_{het}, \sigma_{inf}, x^*, x_0^{st})$ be a vector of all parameter values, lying in the parameter space $\mathscr{C} = \mathbb{N}_{>0}^2 \times \mathbb{R}_{>0}^5$. Each result below specifies its assumptions as a restriction on this parameter space $\mathscr{C}$.

**Result 1.** *Let $b = |x_0^{st} - x^*|$ and let $\sigma_{st}, \sigma_{het}, \sigma_{inf} > 0$. Let $\mathscr{A}^{gen}$ and $\mathscr{A}^I$ be the optimal accuracies when updating about a random member m and encountered member I, respectively. For visible attributes, the following results hold:*
*i)* ***Diminishing returns to contact:*** *Assume $b^2 \geq \frac{2}{3}\sigma_{st}^2$. Ceteris paribus, increasing either n or D increases $\mathscr{A}^{gen}$, with diminishing returns.*
*ii) n* ***versus*** *D* ***trade-off in*** *$\mathscr{A}^{gen}$****:*** *For any fixed $N = nD$, $b^2 \geq \frac{2}{3}\sigma_{st}^2$ guarantees that increasing n strictly increases $\mathscr{A}^{gen}$.*
*iii) n* ***versus*** *D* ***trade-off in*** *$\mathscr{A}^I$****:*** *For any fixed $N = nD$, increasing D strictly increases $\mathscr{A}^I$.*
*iv)* ***Breadth increases responsiveness to encountered members:*** *Let $\bar{x} = \frac{1}{nD}\sum_{i=1}^{n}\sum_{d=1}^{D} x_{id}$. Holding total contact $N = nD$ fixed, the responsiveness of the guess about a stranger to*

*the average observation,*

$$\frac{\partial \widehat{x_{n,D}}^{gen}}{\partial \bar{x}} = 1 - w^{st} = \frac{N\sigma_{st}^2}{N\sigma_{st}^2 + \sigma_{inf}^2 + \frac{N}{n}\sigma_{het}^2},$$

*is strictly increasing in n. This result holds irrespective of the initial error b.*

*Proof.* See Appendix C.4. □

Here is some intuition about the results and the assumptions that support the results: As $n, D$ increase, $w^{st}$ decreases in expressions (5) and (6), thus decreasing the inaccuracy originating in the initial prior—this is a primary channel through which contact operates. Getting additional noisy observations (increasing $n$ or $D$) may be pointless if the initial measure $x_0^{st}$ was highly precise to begin with. Positive returns to $n$ and $D$ are guaranteed when $x_0^{st}$ is sufficiently inaccurate relative to the DM's assumed variance $\sigma_{st}^2$. This is assured by the condition $b^2 \geq \frac{2}{3}\sigma_{st}^2$ in (1.i). The same condition about a sufficiently inaccurate prior guarantees that broader contact beats deeper contact when updating about a random outgroup member (1.ii). Intuitively, even though both $D$ or $n$ can decrease $w^{st}$ in isolation, increasing $D$ at the expense of $n$ raises the covariance between sample points (through $\eta_i$), thus decreasing accuracy. Finally, deeper contact beats broader contact when updating about an encountered member. Intuitively, repeated observations of the same partner help identify that partner's individual-specific component $\eta_I$, whereas observations of other outgroup members do not. Thus, holding total contact fixed, greater depth improves partner-specific learning whenever there is positive heterogeneity across outgroup members.

Part (iv) provides a further testable implication: for the same change in average observed partner attributes, beliefs about outgroup strangers should respond more strongly under broad contact. Section 5.5 examines this prediction using random variation in the pre-treatment productivity of assigned partners.

**Discussion** In the model, misperceptions originate from noisy observations and the prior (stereotypical) parameter $x_0^{st}$ being different from $x^*$, while the process of updating the guess is completely rational. It is plausible to imagine alternative behavior, where the process of information retrieval and processing can be additional sources of inaccuracy. For example, with deep contact, it could be the case that signals about the partner are completely discounted, with motivated beliefs driving the participant to consider their outgroup partner an exception to the rule.

For visible traits, the informational value of a deeper intervention (increasing $D$) hinges on the assumption that subsequent days of interaction with the same individual are meaningfully informative. Such an assumption is only valid in settings where learning is indeed gradual and its precision increases over multiple meetings with the same $i$. But, for visible attributes that can be learned immediately in one meeting, or more generally, in settings where the DM repeatedly observes the first day's observation $x_{i1}$ on every subsequent meeting with $i$, the deep treatment should instead be modelled as only offering $n$ (instead of $nD$) data points. Thus, in that case deeper interventions would offer zero informational gains. We show some evidence for this below when considering effects on productivity beliefs.

**DM as a Bayesian** In Appendix C.5, we show that all of our results are robust to modeling the DM as a Bayesian—one who begins with an explicit prior distribution over the unknown group average and updates it based on his interactions. In particular, we prove that such a Bayesian DM arrives at posteriors centered exactly at the expressions given by Equations 5 and 6. The derivation relies on the standard assumptions of normality of the prior and of all random variables $\eta$ and $\varepsilon$.[26] The Bayesian framework and its derivation are notation-intensive without offering any additional insights, and are therefore relegated to the appendix.

### 4.2 Learning non-visible attributes through friendship

Each interaction with an outgroup member not only provides information about visible attributes, but also allows an opportunity to form a new friendship, which we assume is necessary for learning about non-visible attributes. But, friendship formation requires investing time and (cognitive) effort in seeking commonalities and striking up meaningful conversations with the other person. While the investment requires a one-time cost, forming a friendship with the partner increases the utility of the DM throughout the period of the interaction.

**Investment in friendship** We assume that the individuals, in our environment, are only interested in forming friendship because of the flow utility it provides, and not because of the learning it facilitates. Thus, the decision on whether to invest in a friendship, or more generally, the willingness to interact with the outgroup member, depends on three factors: the one-time cost of investment, the perceived utility of friendship $\hat{x}^i$[27], and the duration of the relationship through which its (relative) benefits can be enjoyed.

[26] Note that in Section 4.1, we made no parametric assumptions about the distributions of the prior or the noise beyond their first and second moments.

[27] This can be different if the partner is an ingroup or outgroup.

At the beginning of the first interaction with outgroup member $i$, the DM chooses whether to pay the one-time investment cost $c > 0$. Let $\widehat{x}^i$ denote the expected utility per interaction from friendship and let $D$ denote the anticipated total number of interactions. The DM invests if

$$D\widehat{x}^i - c \geq 0.$$

Assuming $\widehat{x}^i > 0$, this is equivalent to $D \geq c/\widehat{x}^i = D^*$.

For $D > D^*$, investment in friendship begins from the first interaction, in anticipation of a sufficiently long relationship. This investment facilitates learning about attributes and the sampling logic from Result 1 continues to apply: information learned from a partner can inform beliefs about the wider outgroup, but such generalization is stronger when contact is broader.

Instead, if $D < D^*$, the decision-maker does not invest in the relationship and therefore receives no informative signals about these attributes. In this case there is neither partner-specific learning nor generalized learning. Further, for any fixed amount of total contact $nD$, the requirement $D > D^*$ restricts the breadth of contact, i.e., the number $n$ of distinct individuals with whom such learning can occur. Hence, learning about attributes that require relational investment faces a particularly strong sampling constraint: deeper contact facilitates learning about the individual partner, while the resulting small breadth limits how much that learning carries over to beliefs about the wider outgroup. Thus, generalized learning about such attributes is most likely when contact is comprehensive enough to be sufficiently deep to reveal the attribute and sufficiently broad to provide multiple independent observations.

**Predictions about interactions, friendship, and learning** This simple model leads to the following predictions:

1. For $D^* \in (1,6)$, subjects in the *Outgroup, deep* treatment invest more in the relationship than subjects in *Outgroup, broad*, including from the first day when the treatments differ only in anticipated future interaction.
2. As a result, *Outgroup, deep* is likely to generate closer ties with work partners, and a happier work environment, under the assumption that closer ties promote happiness when interacting.
3. Still, due to the lack of sufficient independent samples, this attitude change generates limited generalization towards the larger outgroup.
4. If $D^* > 1$, *Outgroup, broad* may fail to reveal attributes that require sufficiently prolonged relational investment, despite its greater sampling breadth.

# 5 Empirical results

## 5.1 Daily mood

Assignment to the *Outgroup, broad* condition produced a work environment that was substantially different from the environments workers experienced in the two deep conditions (Table 3). Workers in *Outgroup, broad* were less happy than those in *Outgroup, deep* (column 1, $p = 0.06$), they felt their relationship with their partner was more distant (column 2, $p < 0.01$), and they described themselves as being less close friends (column 3, $p < 0.01$). These less positive moods are also echoed in performance assessments, with subjects consistently ranking both their own and their partner's performance as worse in the broad condition than in the deep condition (columns 4 and 5, both $p < 0.01$). We show in Subsection 5.6 that the negative effect on perceived performance comes despite null effects on objectively measured productivity.

Deep and broad contact differ with respect to the number of interactions a participant has with a given work partner. This difference plausibly affects happiness and partner friendships. But deep and broad contact also differ along a more subtle dimension: prior to the final day, deep participants anticipate future interactions with their current partner, whereas broad participants do not. As we laid out in our theory, this anticipation of future interaction may itself affect present interactions; for example, if I know I will have to interact with my partner in future, I may try harder to avoid friction today, and I may invest more effort in getting to know my partner. To test this mechanism, we compare *Outgroup, deep* and *Outgroup, broad* participants on the first day of the intervention in Appendix Table S6; at this point, the number and depth of interactions with the outgroup is the same for both groups, but the anticipation of future interaction differs. Deep participants are $0.15\sigma$ happier ($p = 0.09$), and report their partner to be $0.17\sigma$ closer as a friend ($p = 0.05$). The mere expectation of future outgroup interaction is enough to change current emotions and interactions.

Our theory predicts that deep contact should lay the foundations for deeper and more friendship-inducing discussions with partners: when you know you will be spending more time with someone, you are more willing to invest in them and steer conversations beyond platitudes. That is what we find. Table 4 examines treatment effects on the content of conversations with partners on the sixth and final day of work (the only day we measured such data). Overall, subjects talk about fewer topics in *Outgroup, broad* than in *Outgroup, deep* (column 7, $p = 0.03$). Participants in *Outgroup, broad* are less likely than those in *Outgroup, deep* to have talked about TV shows, movies, music and books (column 1,

Table 3: Subjects' mood and relationships are worse in *Outgroup, broad*.

| | Outcome ($z$-scores): | | | | |
|---|---|---|---|---|---|
| | Happiness today (1) | Partner relationship (2) | Partner close friend (3) | Own performance (4) | Partner's performance (5) |
| Outgroup, deep | 0.09 | 0.06 | 0.05 | 0.12 | 0.08 |
| | (0.07) | (0.07) | (0.08) | (0.08) | (0.08) |
| Outgroup, broad | -0.00 | -0.26*** | -0.42*** | -0.14* | -0.21*** |
| | (0.07) | (0.07) | (0.07) | (0.08) | (0.08) |
| *N* | 3,997 | 3,998 | 3,997 | 3,998 | 3,997 |
| Outgroup, deep vs. Outgroup, broad $p$-value | 0.06 | <0.01 | <0.01 | <0.01 | <0.01 |

*Notes:* This table reports estimates of treatment effects on subjects' mood, perceptions about the partner, and work self-assessments, as reported in daily surveys taken on each day of the intervention. The survey questions for each column are: (1) Which picture best describes your emotions during today's work? (0 = Very sad face to 4 = Very happy face); (2) Which picture best describes your relationship with your partner today? (0 = self and partner circles apart to 4 = self and partner circles fully overlapping); (3) How close of a friend do you consider your partner to be? (1 = Not at all to 10 = To a great extent); (4) How would you rate your personal performance during today's work? (1 = Very bad to 10 = Excellent); (5) How would you rate your partner's performance during today's work? (1 = Very bad to 10 = Excellent). All outcomes are $z$-scores, centered and standardized according to the mean and standard deviation of the variable in the *Ingroup, deep* condition. Only subjects assigned to work are included in the regressions. We drop surveys for subjects whose partner was absent that day. The unit of analysis is subject-work day. Regressions include randomization strata fixed effects. Robust standard errors, clustered by subject, are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

$p = 0.01$), as well as politics (column 6, $p = 0.01$), hobbies (column 5, $p = 0.04$), and sports (marginally so, column 2, $p = 0.12$). Instead, they appeared to focus on matters of work and personal life—talking about these topics at statistically similar rates to *Outgroup, deep* subjects (columns 3 and 4, respectively).

Taken together, the rotation of work partners gives rise to an atmosphere that is more isolating and less happy than a situation in which mixed-group workers are permanently paired. However, and surprisingly, we find no evidence of such effects on Hindus from being partnered with an outgroup partner rather than an ingroup partner: each point estimate on *Outgroup, deep* in Table 3 is actually positive, although not statistically significant. At endline, *Outgroup, deep* workers also recall being similarly happy during the work-period as *Ingroup, deep* workers (column 1, Table S7), though they are 13 percentage points less likely to accurately recall their partner's name (column 5). Deep outgroup contact does not have negative affective impacts in the moment, perhaps because the anticipation of future interaction mitigates intergroup frictions, but down the line, it is less likely to be remembered.

### 5.2 Social and economic ties

The pattern of effects on mood and closeness during the work period is also reflected in the subsequent social and economic ties of participants. Hindus in the deep contact condition

Table 4: *Outgroup, broad* leads to fewer and less intimate conversations between partners.

| | Topic of conversation: | | | | | | |
|---|---|---|---|---|---|---|---|
| | TV shows/ movies/ music/books (1) | Sports (2) | Work (3) | Personal life (e.g., family relations) (4) | Hobbies (5) | Politics (6) | Number of topics (7) |
| Outgroup, deep | 0.00 | 0.02 | -0.02 | 0.03 | -0.01 | 0.04 | 0.06 |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) | (0.04) | (0.15) |
| Outgroup, broad | -0.09** | -0.05 | 0.00 | 0.06 | -0.09* | -0.03 | -0.20 |
| | (0.04) | (0.05) | (0.05) | (0.05) | (0.05) | (0.03) | (0.14) |
| *N* | 699 | 699 | 699 | 699 | 699 | 699 | 699 |
| Ingroup, deep, outcome mean | 0.27 | 0.66 | 0.76 | 0.66 | 0.43 | 0.12 | 2.91 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.01 | 0.12 | 0.58 | 0.35 | 0.04 | 0.01 | 0.03 |

*Notes:* This table reports estimates of treatment effects on the probability of particular conversation topics being discussed, and on the total number of topics discussed, on day 6 of the intervention. Outcomes in columns 1–6 are binary variables based on responses to the survey question: "Today, what topics did you discuss with your partner? (Select all that apply)." The outcome in column 7 is the sum of those binary variables. Only subjects assigned to work are included in the regressions. The unit of analysis is the subject. Regressions include randomization strata fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

build similar ties with ingroup and outgroup partners. At endline, 59% of *Ingroup, deep* participants correctly report their partner's phone number, with no statistically significant difference relative to *Outgroup, deep* participants (column 1, Table 5). Participants have similar post-experiment phone contact with ingroup and outgroup partners (column 2), though there is suggestive evidence of less in-person contact with outgroup partners ($p = 0.14$, column 3). In the economic domain, deep participants are, if anything, more likely to have received help finding work from, or to have borrowed from or lent money to, outgroup partners (columns 4 and 5).

While having an outgroup (rather than ingroup) partner does not hurt subsequent social ties, interacting only for one day rather than six does. *Outgroup, broad* participants are 41 percentage points less likely to correctly report a given partner's phone number (they are asked about their partners on days 1, 4, and 6) than *Outgroup, deep* participants ($p < 0.01$, Table 5, column 1). Broad participants are 39 percentage points less likely to have had recent phone contact with a given partner, and 10 percentage points less likely to have met with them in-person (both $p < 0.01$, columns 2 and 3). The weaker ties for broad partners also have economic implications: *Outgroup, broad* participants receive less help finding work from a given partner,[28] and are less likely to borrow from or lend to a given partner than *Outgroup, deep* participants (columns 4 and 5).

Table 5 shows that participants have on average weaker ties with one-day partners than with six-day partners. Nevertheless, the network benefits of broad contact may dominate those of deep contact; for example, six weak ties may facilitate more information diffusion

[28]This result parallels the non-experimental finding of Gee et al. (2017) that an additional strong tie on Facebook is more helpful for job transitions than an additional weak tie.

Table 5: *Outgroup, broad* reduces in-person and remote contact with work partners, and job and financial assistance.

| | Outcome: | | | | |
|---|---|---|---|---|---|
| | Knows partner's phone number (1) | Phone contact (2) | Food, family, in-person contact (3) | Work help (4) | Money help (5) |
| Outgroup, deep | -0.05 | -0.01 | -0.06 | 0.05* | 0.02 |
| | (0.05) | (0.05) | (0.04) | (0.03) | (0.02) |
| Outgroup, broad | -0.46*** | -0.40*** | -0.16*** | -0.06** | -0.01 |
| | (0.04) | (0.04) | (0.04) | (0.02) | (0.01) |
| *N* | 1,207 | 1,207 | 1,207 | 1,207 | 1,207 |
| Ingroup, deep, outcome mean | 0.59 | 0.54 | 0.28 | 0.10 | 0.02 |
| Outgroup, deep vs. Outgroup, broad *p*-value | <0.01 | <0.01 | <0.01 | <0.01 | 0.01 |

*Notes:* This table presents treatment effect estimates on recent interactions with work partners. Only subjects assigned to work are included in the regressions. Regressions include randomization strata fixed effects. The unit of analysis is the subject/partner pair. In *Ingroup, deep* and *Outgroup, deep*, the single partner was asked about. In *Outgroup, broad*, partners from Day 1, 4, and 6 were asked about. The outcome in column 1 is a binary variable for whether the subject reported having the phone number of the partner, with the reported phone number matching a phone number provided by the partner at baseline. The survey question for columns 2 and 3 was: "Please tell me which of the following types of contact (if any) you have had with [partner name] over the past week." The outcome in column 2 is a binary variable for whether the subject reported "Called or sent message via phone." The outcome in column 3 is a binary variable for whether the subject reported any of the following types of contact: "Seen in person," "Taken food or drink together," or "Spent time with each other's families." The outcome in column 4 is a binary variable for whether the subject reported that the partner helped them to find paid work in any way since they worked together at the factory. The outcome in column 5 is a binary variable for whether the subject reported ever borrowing from or lending money to the partner. Robust standard errors, clustered by subject, are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

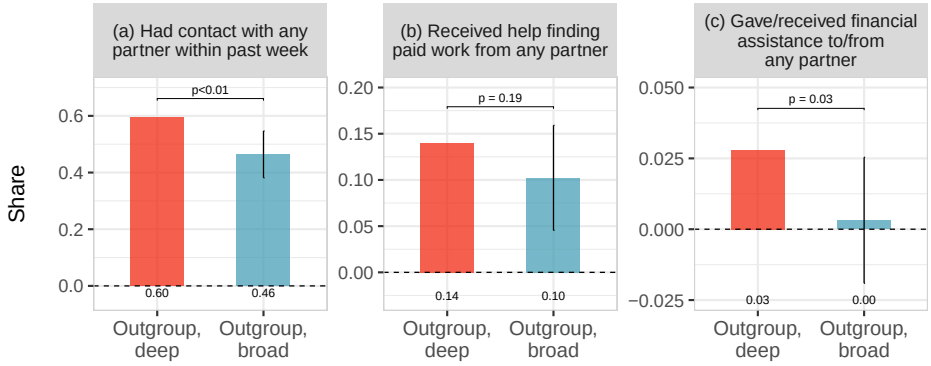
about economic opportunities than one strong tie. Alternatively, the strongest of the six social ties may be stronger than the average tie with a deep partner, because broad contact allows a participant to select their closest friend from among six options. We test for these ideas in Figure 3 by using survey questions that pertain to *all* work partners of a given participant, rather than just one.

Though *Outgroup, broad* participants have more past partners that they can contact, *Outgroup, deep* participants are 14 percentage points more likely to have had contact with any work partner over the last week ($p < 0.01$, Panel (a), Figure 3). One strong tie is then delivering more subsequent social contact than six weak ties. The same gap is reflected in economic interactions: *Outgroup, broad* participants are four percentage points less likely than *Outgroup, deep* participants to receive help finding work ($p = 0.19$, Panel (b)), and three percentage points less likely to have borrowed from or lent money to work partners ($p = 0.03$, Panel (c)).[29]

The findings in Figure 3 speak against one particular version of the strength of weak ties theory (Granovetter 1973), since in our setting, one strong outgroup tie dominates six weak

[29]Appendix Table S8 shows that there is no consistent pattern of heterogeneity in these effects by participant religion.

Figure 3: Subjects have weaker ties with former work partners in *Outgroup, broad*.




*Notes:* This figure presents the estimated treatment effects on respondents' relationship with former work partners during the endline survey. The sample includes subjects assigned to *Outgroup, deep* or *Outgroup, broad* only. The *Outgroup, deep* bar shows the average for subjects in the *Outgroup, deep* condition. The other bar adds treatment effects estimates to that bar, estimated from a regression of the outcome on the *Outgroup, broad* indicator and randomization strata fixed effects. The 95% confidence intervals and $p$-values for the difference are based on robust standard errors. Numbers under the dashed horizontal lines are levels. The outcomes are: (a) A binary variable for whether the subject reports having had contact (in-person or by phone) with (any of) their work partner(s) within the last week; (b) Whether the subject reports having received help finding paid work from (any of) their work partner(s); (c) Whether the subject reports having borrowed from, or lent to, (any of) their work partner(s).

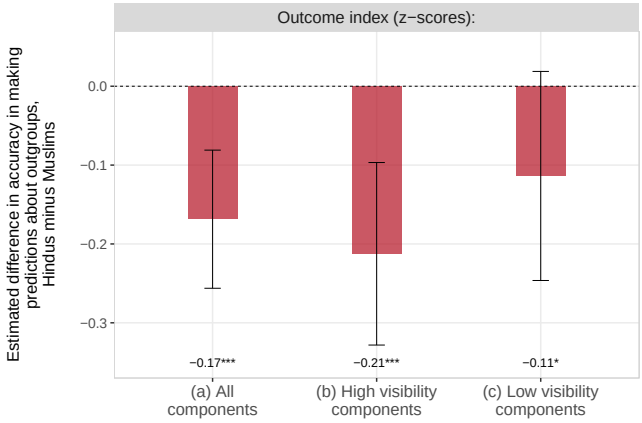
outgroup ties. However, the findings in Table 5 give some support to an alternative version of the theory. In particular, the results suggest that one strong outgroup tie might dominate one strong ingroup tie, with deep participants somewhat more likely to get help finding work from, and to lend to or borrow from, an outgroup partner than from an ingroup partner. This result is consistent with the idea that information on economic opportunities from an additional ingroup contact is more highly correlated with information already available from a participant's existing social network, and thus more redundant than information available from outgroup members.

### 5.3 Beliefs about outgroup strangers

Our results so far show that participants build similar ties with fixed ingroup and outgroup partners, and deep contact leads to more positive work experiences as well as stronger subsequent ties with partners than broad contact. In this sub-section we explore the theoretical prediction that broad contact leads to greater generalized belief updating.

Before turning to the estimated treatment effects, we first show that in the control group, Muslims are more accurate in predicting the characteristics of Hindu strangers than Hindus are about Muslim strangers (Figure 4). We report differences in the average accuracy (between Hindus and Muslims) in predictions made about outgroup members using the belief accuracy index described in Section 3. Muslims are more accurate overall, and for the high and low visibility attributes separately. This makes sense given that Muslims—the minority—are

Figure 4: Control group Hindus make less accurate predictions about the outgroup than Muslims.




*Notes:* This figure shows the difference in average accuracy, in standard deviations, between Hindus and Muslims in making predictions about outgroups, using only control group participants. In column (a), all nine components on which subjects made guesses about the holdout sample (or about factory workers for productivity guesses), are standardized and averaged in a single index. Column (b) uses only the five high-visibility components: productivity, bag, identity, children, and sanitizer. Column (c) uses only the remaining four low-visibility components: coin, risk, movie, and dictator. *p<0.1; **p<0.05; ***p<0.01.

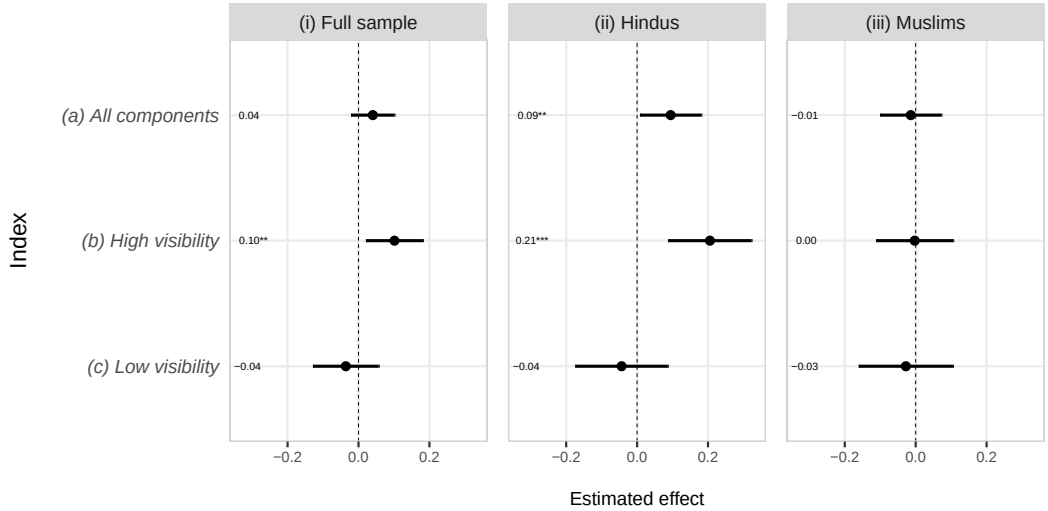
exposed to many more Hindus—the majority—at work and in their residential wards than vice versa (Appendix Figure S7), meaning they have less scope for updating. This is also consistent with our theoretical framework. Furthermore, Muslims are no more accurate about guessing the characteristics of *ingroups* (Appendix Figure S9), ruling out the alternative explanation of Muslim subjects being more attentive in general when answering belief questions.

**Generalized belief updating** Figure 5 reports treatment effects on generalized belief updating.[30] We show effects for *Outgroup, broad* relative to *Outgroup, deep*, the two treatment conditions that involve outgroup exposure. The evidence aligns with the theoretical prediction of broad contact being more effective for generalized belief updating, although even broad contact has limited effects.

For the full sample, broad contact has an insignificant +0.04$\sigma$ effect on generalized belief accuracy ($p = 0.20$, row (a), column (i)). However, the effects differ for the majority and minority groups: broad contact increases the belief accuracy of Hindus (0.09$\sigma$, $p$ = 0.03, row (a)), but not Muslims (-0.01$\sigma$, insignificant), in keeping with the latter's higher accuracy about outgroup characteristics in the control group, and our theoretical prediction of diminishing returns. Consistent with that prediction, we also find that broad contact improves accuracy less for participants who already live among the outgroup—each 20-percentage-point increase in the outgroup share in their residence ward lowers the

[30]We present results for the individual components in Appendix Figure S11.

Figure 5: *Outgroup, broad* increases
prediction accuracy about the more visible preferences and behaviors of outgroup strangers.




*Notes:* This figure shows estimated treatment effects (in standard deviations) on the accuracy of predictions made about outgroups in the holdout sample. The unit of analysis is the subject. The sample includes subjects in the *Outgroup, deep* and *Outgroup, broad* conditions. All outcomes are $z$-scores, centered and standardized according to the variable mean and standard deviation in *Outgroup, deep*. The outcome for panels in row (a) is the mean of the nine components measuring prediction accuracy, as described in the main text. The outcome for panels in row (b) is the mean of the five components: productivity, bag, identity, children, and sanitizer. The outcome for panels in row (c) is the mean of the remaining four components: coin, risk, movie, and dictator. Regressions include randomization strata fixed effects. The 95% confidence intervals and $p$-values for the differences are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

broad-versus-deep effect by $0.07\sigma$ ($p = 0.03$).

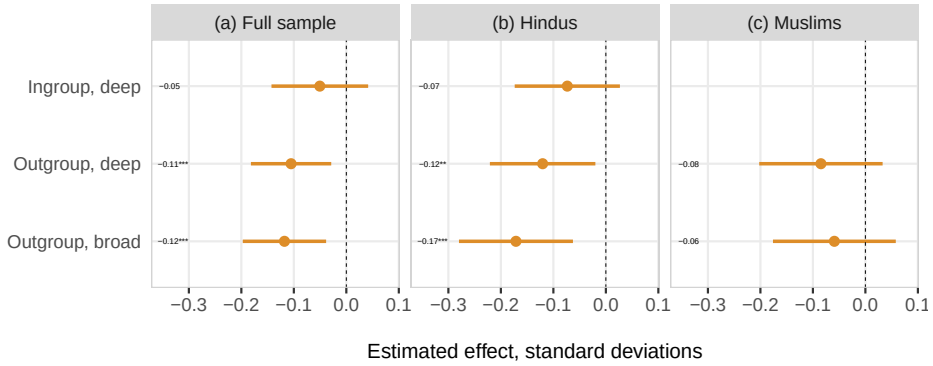
The effects of broad contact are driven by the high-visibility attributes (productivity, religious identity, fertility preferences, and hygiene), with broad contact increasing the accuracy of high-visibility beliefs by $0.1\sigma$ ($p = 0.01$, full sample of Hindus and Muslims), and null effects on the low-visibility attributes (rows (b) and (c), Figure 5).

In contrast to the effects of broad relative to deep contact, we estimate null effects of *Outgroup, deep* contact relative to the pure control group (Appendix Figure S10).[31] The distinction between broad and deep contact is then first order: it is not just that broad contact improves marginally on the generalized belief accuracy already caused by deep contact. With deep contact, we see no evidence for belief generalization improving accuracy at all.

Our endline survey instrument elicited a parallel set of beliefs about a randomly selected sample of ingroup individuals (that is, about a Hindu holdout individual for Hindu subjects and a Muslim holdout individual for Muslim subjects). This allows us to create an index

[31]We use the comparison with pure control here, rather than *Ingroup, deep*, given that only Hindus are assigned to *Ingroup, deep*. The comparison with pure control then ensures that Appendix Figure S10 is fully comparable with Figure 5.

Figure 6: Both *deep* and *broad* contact with outgroups lead to more similar guesses about Hindus and Muslims.




*Notes:* This figure estimates treatment effects on the absolute difference between subjects' guesses about Hindus and Muslims in the holdout sample (or in another wave, in the case of productivity). We did not ask participants to predict the type of bag picked by an ingroup stranger, so that outcome is not part of this index. The unit of analysis is the subject. Estimates are based on a regression of this index on the treatment indicators and randomization strata fixed effects. 95% Confidence intervals are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

of prediction accuracy regarding ingroups that precisely parallels the one used for outgroups. Our theoretical expectation is that there should be no difference between the *Outgroup, deep* and *Outgroup, broad* groups in terms of the accuracy of predictions made about ingroup strangers, since there was no additional contact with ingroups. As predicted, we estimate precise null effects for this placebo outcome when employing the full sample (Panel (i), Appendix Figure S12) or religious subgroups (Panels (ii) and (iii)).

**Any contact has a humanizing effect** In Figure 6, we investigate a second aspect of the guesses subjects were asked to make; namely, the degree to which our treatments induce subjects to offer *more similar* predictions about Hindus and Muslims in the holdout sample. Using each of the outcomes for which subjects made predictions about both ingroup and outgroup strangers, we create a standardized measure to capture the difference between the predictions made for the ingroup versus outgroup member. Subjects in both the *Outgroup, deep* and *Outgroup, broad* conditions made guesses about holdout Hindus and Muslims that are more similar, compared to subjects in the control group ($p < 0.01$ for both cases). As expected, the effects for those in *Ingroup, deep* are indistinguishable from zero. Regardless of accuracy then, contact with either one or many outgroup members has a humanizing effect, reducing the tendency to see outgroups and ingroups as fundamentally dissimilar. Once more, this effect is driven by majority-group Hindus, consistent with their lower baseline exposure to Muslims.

Table 6: *Outgroup, deep* and *Outgroup, broad* both increase prediction accuracy about partner productivity.

| | Productivity prediction accuracy: | |
|---|---|---|
| | (1) | (2) |
| Mixed team | 0.59** | |
| | (0.28) | |
| Outgroup, deep | | 0.57* |
| | | (0.30) |
| Outgroup, broad | | 0.61* |
| | | (0.32) |
| | | |
| *N* | 726 | 726 |
| Control, outcome mean | -4.11 | -4.11 |
| Outgroup, deep vs. Outgroup, broad *p*-value | | 0.89 |

*Notes:* This table shows treatment effects on the accuracy of beliefs about the productivity of individual partners. The outcome variable is the negative absolute difference between a study participant's prediction about their outgroup partner's productivity on the final day of the lead-in period (i.e., before they worked together) and the actual number of bags produced by the partner on the day. For participants in the *Outgroup, broad* group, we randomly selected their partner on day 1 or day 6 of the intervention for this prediction. The control group is the omitted category in this regression. To obtain comparable predictions from the control group, we matched them with outgroup workers in the factory (pseudo-partners) from a different wave of the experiment. Regressions include randomization strata fixed effects. Robust standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

### 5.4 Beliefs about outgroup partners

In order to change one's beliefs about outgroup strangers, one might expect it to be necessary to first alter beliefs about the individuals one has direct contact with. The *Outgroup, broad* treatment led Hindus to have more accurate beliefs about Muslim strangers, but did it also make them more accurate in their beliefs about their Muslim work partners? Further, do subjects in *Outgroup, deep* make more accurate predictions about their work partners than subjects in *Outgroup, broad*, as our theoretical model predicts? To answer these questions, we use measures of workers' beliefs about work partners, rather than strangers. We caveat that these belief measures are more limited than our measures of beliefs about holdout sample members, as (i) we do not incentivize these measures, to avoid gaming in which a worker colludes with their partner to learn the correct answer, and (ii) to minimize survey length, we measure only five of the nine belief domains, and we only have an accuracy benchmark for four of the five.[32]

We first consider beliefs about productivity. We asked treated participants to predict the lead-in period productivity of their partner from the six-day intervention period. For

[32]Moreover, eliciting guesses about partners is conceptually problematic for religious identity and the dictator game (both measured at endline), since contact with outgroups may be expected to shift a partner's traits, so respondents are asked to predict a moving target.

those in the *Outgroup, broad* treatment, we randomly selected either their partner on day 1 or day 6 to be guessed about. We matched control group participants with a randomly chosen ingroup and outgroup individual (pseudo-partners) from the factory and obtained similar predictions for comparison (we exclude those in *Ingroup, deep* from this analysis because they only have ingroup partners). Table 6 illustrates that both the *Outgroup, broad* and *Outgroup, deep* treatments led participants to provide more accurate predictions about their partner relative to the control group. Contrary to the model, the effect sizes are very similar—suggesting that five additional days of seeing the partner at work did not deliver additional informational gains. Workers were then able to update about their partners fairly quickly, at least for the most visible characteristic in this experiment, productivity.

We report parallel effects for all other measured dimensions in Table S11. For beliefs about desired fertility, religious identity, and dictator game behavior, we report effects on both the level of the belief, as well as the accuracy. For beliefs about coin toss behavior, we only report effects on the level, as we lack the true benchmark. Both broad and deep outgroup contact shift beliefs about partners, causing Hindu participants to update their beliefs downward about the desired fertility of their Muslim work partner (panel A, column 1), and Muslims to update downwards about the national identification of their Hindu partners (panel B, column 3), and upward about their generosity (column 5).

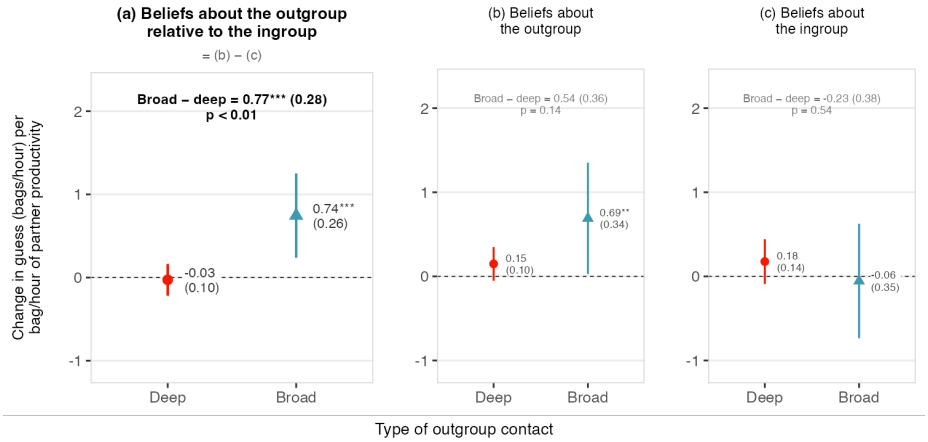
While contact shifts beliefs, the prediction of the model is that deep contact should increase accuracy more than broad contact. In reality, we never reject the null that deep and broad contact have the same effect on accuracy (columns 2, 4, and 6), nor do we reject the null that they have the same effect on the levels of beliefs. That said, when combining the three accuracy measures into an index, we find suggestive evidence of larger accuracy effects of deep contact for Hindus ($p = 0.14$).

Summarizing, both types of contact cause belief updating about the individual interacted with, though we do not see clear evidence for the model prediction that deep contact should facilitate more accurate beliefs about the individual than broad contact. The preference dimensions we study tend to be learned quickly, over one day, or not at all.

### 5.5 Do beliefs about strangers track the partners subjects met?

Our results thus far compare average beliefs across treatment arms. Within each arm, the partners with which each subject worked were determined by randomization. Consequently, the pre-treatment productivity of those partners, measured during the lead-in period, varies

Figure 7: Beliefs about outgroup strangers track the productivity of partners met under broad contact.




*Notes:* Each point is the slope of an endline productivity guess (bags per hour) on the lead-in productivity of the subject's assigned partners, measured relative to the average productivity of the pool from which partners were drawn (all subjects of the partners' religion assigned to work in the same wave and shift). For *Outgroup, deep*, this is the single partner's productivity; for *Outgroup, broad*, it is the average across the six assigned partners. Panel (a) uses the difference between the guesses about an outgroup and an ingroup stranger; panels (b) and (c) show its two components. Regressions control for the pool average, the subject's own lead-in productivity, and randomization strata fixed effects. The sample includes the 505 subjects in *Outgroup, deep* and *Outgroup, broad*. 95% confidence intervals are based on heteroskedasticity-robust standard errors.

randomly across subjects. If broad contact promotes generalization, beliefs about outgroup strangers should respond more strongly to partners' productivity under broad than under deep contact (Result 1.(iv) of our model).

We regress subjects' endline productivity guesses on the lead-in period productivity of their assigned partners—their single partner in the deep arm, or the average of their six partners in the broad arm—interacted with the *Outgroup, broad* treatment dummy.[33] Having a productive outgroup partner can teach a respondent that the task may be easy and people might generally be productive at it, which should raise guesses about both ingroup and outgroup strangers. But it could also teach a subject that members of the outgroup are productive, which should raise guesses only about the outgroup. Our main outcome is therefore the difference between the guesses about an outgroup and an ingroup stranger, which isolates the outgroup-specific component.

Figure 7 presents the results. Under deep contact, a more productive partner raises guesses

[33] Specifically, we estimate $y_i = \psi_s + \beta_1 \text{PeerProductivity}_i + \beta_2 (\text{PeerProductivity}_i \times \textit{Outgroup, broad}_i) + \gamma \textit{Outgroup, broad}_i + \delta \text{Pool}_i + \theta \text{Own}_i + \varepsilon_i$ for subjects in the two outgroup conditions, where $\psi_s$ are randomization strata fixed effects. $\text{PeerProductivity}_i$ is partners' lead-in productivity minus $\text{Pool}_i$, the average lead-in productivity of all randomized subjects of the partners' religion in the same wave and shift, so that $\text{PeerProductivity}_i$ reflects only the luck of the draw. $\text{Own}_i$ is the subject's own lead-in productivity. As expected under random assignment, $\text{PeerProductivity}_i$ is unrelated to subjects' own productivity and baseline productivity beliefs (all $p > 0.3$).

about outgroup and ingroup strangers by similar amounts (0.15 and 0.18 bags per hour, per bag per hour of partner productivity; panels (b) and (c)), so the outgroup-specific component is essentially zero ($-0.03$; panel (a)). Under broad contact, partners' productivity shifts guesses about outgroup strangers (0.69), but not about ingroups ($-0.06$), so the outgroup-specific component is 0.74. The difference between broad and deep in the outgroup-specific component is 0.77 ($p < 0.01$). The results are similar when we measure partners' productivity on the third lead-in day only (broad minus deep: 0.50; $p = 0.02$), and when we use the partners subjects actually worked with rather than those assigned (0.67, $p = 0.01$, Appendix Table S12).

These results connect to the two sets of belief findings. Section 5.4 shows that subjects in both arms learn about their own partners' productivity equally quickly. The results in this section show that only broad subjects generalize those beliefs to the outgroup as a whole. The difference between the two treatments thus lies less in what individuals learn about their partners than in whether the belief updates generalize.

### 5.6 Partner rotation and productivity

Religious diversity can negatively impact team production through taste-based discrimination (Hjort, 2014), weakened reputation concerns around outgroup members (Afridi et al., 2020), and biased beliefs about outgroup productivity in the presence of complementarities (Ghosh, 2025).[34] These mechanisms behind negative effects may be more pronounced when outgroup work partners change daily—the limited opportunity to form longer-term relationships with outgroup partners may make it harder to reduce these frictions over time. It may then be that the generalized belief updating in *Outgroup, broad* comes at the cost of productivity losses.

Appendix Figure S6 shows that the average number of big bags produced by teams increased over the course of the intervention, despite the lack of piece-rate incentives, with most of the learning-by-doing occurring in the first three days. However, neither form of Hindu-Muslim mixing has a significant effect on team productivity or quality, as shown in regressions controlling for individual-level productivity during the lead-in period (Table 7). The surprising absence of productivity effects could be due to the straightforward nature of the bag-making task, and the close proximity to the partner in the booth, allowing for quick belief updating regarding the partner's productivity shortly after the start of the day's work. Indeed, our findings in Table 6 reveal that one day of working with an outgroup worker improved the accuracy of productivity beliefs by just as much as six days of working together.

[34] Indeed, we find that prior to randomization, Hindus perceive Hindu workers to be more productive than Muslim workers, whereas Muslims do not display such ingroup bias (Appendix Figure S5).

Table 7: Neither type of contact affects productivity or quality.

| | Small bags | | | Big bags | | |
|---|---|---|---|---|---|---|
| | Number of small bags (1) | Small bag quality (2) | Number of small bags × quality (3) | Number of big bags (4) | Big bag quality (5) | Number of big bags × quality (6) |
| Outgroup, deep | 0.98 | 0.02 | 8.35 | 0.44 | -0.06 | 2.58 |
| | (0.77) | (0.08) | (6.76) | (0.40) | (0.07) | (3.35) |
| Outgroup, broad | 0.68 | 0.03 | 6.49 | 0.28 | -0.04 | 2.12 |
| | (0.71) | (0.07) | (6.31) | (0.38) | (0.07) | (3.19) |
| *N* | 1,989 | 1,988 | 1,988 | 1,989 | 1,987 | 1,987 |
| Ingroup, deep mean | 19.99 | 8.26 | 166.73 | 10.05 | 7.94 | 80.54 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.65 | 0.83 | 0.75 | 0.63 | 0.71 | 0.87 |

*Notes:* This table reports the effects of treatments on production. Each observation is a team (pair) on a given day. We drop cases when only one person in the pair was present on the day. Columns 1–3 report effects on small bags and columns 4–6 on big bags. Small bag quality is the average of the quality scores for the small bags made by each of the two partners. We note that the effects on the quantity of small and big bags are mechanically related given the requirement for pairs to produce two small bags for each big bag they made. Columns 1 and 4 control for the average number of bags produced by both partners during the lead-in period. Columns 2 and 5 control for the average quality of bags produced by both partners during the lead-in period. Columns 3 and 6 control for the average number of bags produced by both partners during the lead-in period multiplied by the average quality of bags produced by both partners during the lead-in period. The regressions include wave × day × shift fixed effects (recall that the religion × wave denotes the randomization strata). Standard errors are clustered at the team-level. *p<0.1; **p<0.05; ***p<0.01.

## 5.7 Generalized intergroup behaviors

We have demonstrated that deep contact dominates broad contact for creating subsequent social and economic ties, whereas broad dominates deep at correcting misperceptions about the outgroup, and neither type of contact has productivity costs. Does either type of contact change behaviors toward outgroup strangers, as per the claims of the contact hypothesis (Pettigrew and Tropp, 2006)? Or do effects on social ties with outgroup partners fail to generalize to the outgroup as a whole, as has been common in pre-registered intergroup contact experiments (Lowe, 2025)? We find evidence more consistent with the latter account in Table 8.

First, although we have shown that roughly one in two participants remain in contact with outgroup partners (panel (a), Figure 3), this does not lead to additional close friendships with outgroup members, whether the contact was broad or deep (column 1, Table 8).[35] This null result on close friendships is in stark contrast to the large, positive, and enduring effects of youth camps on Hindu-Muslim friendships we found in the same municipality (Ghosh et al., 2025). It may be that intergroup friendships are more difficult to forge in adulthood (the average age in the current study is 22, while in Ghosh et al. (2025) it was 15), or that the work environment is less conducive to friendship formation than the recreational camp environment. It is also possible that the higher average number of outgroup participants interacted with at

[35]The participant's five reported closest friends could include factory partners; these friendships would not count as generalization. Nevertheless, even with factory partners included, we do not find effects on outgroup friendships.

Table 8: Neither type of contact leads to generalized changes to intergroup behavior.

| | Outcome ($z$-scores): | | | |
|---|---|---|---|---|
| | Outgroup friends (1) | Reservation wage (2) | Dictator giving (3) | Replied WhatsApp (4) |
| Ingroup, deep | -0.08 | 0.02 | 0.12 | 0.33*** |
| | (0.07) | (0.12) | (0.11) | (0.11) |
| Outgroup, deep | -0.08 | -0.01 | 0.20** | 0.24** |
| | (0.08) | (0.10) | (0.09) | (0.10) |
| Outgroup, broad | -0.04 | -0.08 | 0.14 | 0.26*** |
| | (0.08) | (0.10) | (0.09) | (0.10) |
| *N* | 918 | 918 | 918 | 901 |
| Control, raw outcome mean | 0.20 | 276.36 | 23.25 | 0.28 |
| Outgroup, deep vs. Outgroup, broad $p$-value | 0.60 | 0.47 | 0.47 | 0.83 |

*Notes:* This table presents treatment effect estimates on behaviors toward the wider outgroup. The unit of analysis is the subject. The outcomes for each column are: (1) The number of friends with outgroup-sounding names among subjects' five closest friends; (2) The lowest wage that the subject would be willing to accept to work with a stranger with an outgroup-sounding name; (3) The amount donated to a stranger with an outgroup-sounding name in a dictator game (out of Rs. 100); (4) Whether the subject replied to a WhatsApp message from an outgroup stranger associated with the factory. All outcomes are $z$-scores, centered and standardized according to the mean and standard deviation of the variable in the control group. We use OLS regressions in columns 1, 3, and 4. Column 2 uses a tobit regression, with censoring at the standardized values of the minimum and maximum wage offers made. All regressions include randomization strata fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

the youth camps, and also for longer hours, contributed to the effects on friendships observed there—i.e. unlike the work setting, the campers had both broad and deep outgroup contact.

Second, we do not see evidence of either type of intergroup contact leading to an increase in willingness to interact, as measured by incentivized willingness to work with an outgroup stranger (column 2, Table 8), or in altruism, as measured by a dictator game with an outgroup stranger (column 3). Participants in *Outgroup, deep* do give significantly more in the dictator game than pure control participants, but this difference appears to be driven more by the effect of being assigned to work, and its resultant income gains, than by contact per se.

Third, we estimate effects on civility toward the outgroup by whether the participant responded to a WhatsApp message sent by an outgroup stranger, roughly one month following the final endline. Those in the treated conditions were $0.24\sigma$ to $0.33\sigma$ more likely to respond compared to the control group; however, there were no significant differences between the treatment groups. We interpret the higher response rate in the treated groups as a sign of increased responsiveness to potential future work opportunities, with the treatment group potentially being more eager to engage, having worked in the factory for longer. The similar effects of both *Outgroup, deep* and *Outgroup, broad* suggest that despite deep exposure fostering stronger social ties with partners, it does not influence behavior toward outgroup strangers any more than broad exposure does.

Our theoretical model does not explicitly make predictions about the intergroup behaviors discussed here. However, the model claims that for generalized belief updating about non-visible attributes we need both broad *and* deep contact. One rationalization of our null effects on generalized behavior would then be that generalized behavioral change requires updating about non-visible attributes. Future work could test this through interventions that combine both broad and deep contact.[36]

## 6 Conclusion

Martin Luther King wrote, "I am convinced that men hate each other because they fear each other. They fear each other because they don't know each other."[37] This study addresses a critical issue: how to bridge deep divides in polarized societies by understanding how outgroups may come to know one another better. We present the results of a randomized experiment examining the effects of short-duration exposure to many outgroup members (broad contact) versus longer-duration exposure to just one outgroup member (deep contact). We find that broad contact leads to more generalized belief updating, while deep contact leads to stronger intergroup ties. Participants in both conditions learn about their partners, but only those with broad contact carry this over to beliefs about the outgroup as a whole.

Our findings speak to the question of how to optimize intergroup contact interventions. Decisions about how to structure cross-group interactions exist in everyday life—in universities, factories, and team sports, as well as in political parties and parliaments. Policymakers might harness the lessons gleaned from this study by selecting the type of contact that best fits with their objective function, perhaps placing different weights on belief correction and tie creation. Alternatively, it may be that an intervention with a mix of broad and deep contact could deliver both belief updating and strong intergroup ties. Future work could test this. For example, one could imagine a school in which a student has a fixed outgroup partner for science classes, and rotating outgroup partners for mathematics classes.

An additional line of research could consider how broad contact itself could be designed to maximize belief updating. In particular, unlike our factory environment, the tasks to be carried out in pairs could be chosen to be maximally informative along dimensions that the policymaker cares about. As an example, if the most prevalent stereotype is that outgroup

[36] Beyond intergroup behaviors, we also estimate null effects of both types of intergroup contact on broader political measures: self-reported national identity, ranking of democracy as a political system, and feeling-thermometer warmth toward Narendra Modi (Appendix Table S10).

[37] "Dr. Martin Luther King's visit to Cornell College," *Cornell College News Center*, shorturl.at/28iu3.

members cheat, a piece-rate production task could be designed in which workers self-report their productivity, without these reports being checked. Alternatively, conversations could be structured to be maximally revealing of the traits of outgroup members. Psychologists have already developed a version of this, the "fast friends" procedure, in which a pair of participants ask each other increasingly personal questions (Aron et al. 1997). This procedure has been relatively neglected by economists (though see Clochard et al. (2023) and Clochard (2025) for exceptions).

At the same time, our theory and empirics sound a pessimistic note about the high barriers to achieving generalized behavioral change toward outgroups—and thus the difficulty of prejudice reduction writ large. Wider sampling (through broad contact) yields updating about outgroups, but only regarding easily visible outgroup characteristics; friendship (forged through deep contact) yields stronger bonds with individual outgroup members, but without any generalization. Viewed in this light, our study may help explain why generalization is seldom observed in the intergroup contact literature, particularly given that existing contact interventions are more often deep and less often broad—for example, half of the contact interventions studied in Lowe (2025) involved meeting only one outgroup member.

The final takeaway is that majority and minority groups have distinct responses to contact. In our study, contact does not improve the minority group's accuracy about the majority—because, we suggest, minorities have more everyday interactions with the majority than vice versa, and thus have more accurate beliefs and less scope to learn to begin with. Except in the most segregated settings, this may be a general feature of life as a minority. An implication is that the goal of correcting misperceptions may be of weaker relevance for minority-targeted programming.

## References

**Afridi, Farzana, Amrita Dhillon, and Swati Sharma**, "The Ties that Bind Us: Social Networks and Productivity in the Factory," *CEPR Discussion Paper No. DP14687*, 2020.

**Alesina, Alberto and Eliana La Ferrara**, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

**Allport, Gordon**, *The Nature of Prejudice*, Garden City, NJ: Anchor, 1954.

**Amir, Yehuda**, "The Role of Intergroup Contact in Change of Prejudice and Ethnic Relations," in Phyllis A. Katz, ed., *Towards the Elimination of Racism*, New York: Pergamon, 1976, pp. 245–308.

**Anderson, Michael L.**, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

**Aron, Arthur, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator**, "The experimental generation of interpersonal closeness: A procedure and some preliminary findings," *Personality and social psychology bulletin*, 1997, *23* (4), 363–377.

**Arrow, Kenneth**, "The Theory of Discrimination," *Discrimination in Labor Markets*, 1973, *3* (10), 3–33.

**Bagues, Manuel and Christopher Roth**, "Interregional Contact and the Formation of a Shared Identity," *American Economic Journal: Economic Policy*, 2023, *15* (3), 322–350.

**Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson**, "The diffusion of microfinance," *Science*, 2013, *341* (6144), 1236498.

**_ , Emily Breza, Arun G Chandrasekhar, and Markus Mobius**, "Naive learning with uninformed agents," *American Economic Review*, 2021, *111* (11), 3540–3574.

**_ , Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan**, "Labor Market Discrimination in Delhi: Evidence from a Field Experiment," *Journal of Comparative Economics*, 2009, *37* (1), 14–27.

**Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli**, "Adaptive Linear Step-up Procedures That Control the False Discovery Rate," *Biometrika*, 2006, *93* (3), 491–507.

**Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg**, "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 2019, *109* (10), 3395–3436.

**Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope**, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2025, *107* (3), 605–620.

**Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles**, "Empathy or Antipathy? The Impact of Diversity," *American Economic Review*, December 2006, *96* (5), 1890–1905.

**Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer**, "Stereotypes," *Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

**Bracic, Ana**, *Breaking the Exclusion Cycle: How to Promote Cooperation between Majority and Minority Ethnic Groups*, Oxford: Oxford University Press, 2020.

**Brown, Rupert, James Vivian, and Miles Hewstone**, "Changing attitudes through intergroup contact: The effects of group membership salience," *European Journal of Social Psychology*, 1999, *29* (5-6), 741–764.

**Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott**, "Misperceived social norms: Women working outside the home in saudi arabia," *American economic review*, 2020, *110* (10), 2997–3029.

_ **and David Yang**, "Misperceptions about Others," *Annual Review of Economics*, 2022, *14* (1), 425–452.

_ **, Thomas Chaney, Tarek A. Hassan, and Aakaash Rao**, "The Immigrant Next Door," *American Economic Review*, February 2024, *114* (2), 348–84.

**Camargo, Braz, Ralph Stinebrickner, and Todd Stinebrickner**, "Interracial Friendships in College," *Journal of Labor Economics*, October 2010, *28* (4), 861–892.

**Chakraborty, Anujit, Arkadev Ghosh, Matt Lowe, and Gareth Nellis**, "Identity Uncertainty," Technical Report, National Bureau of Economic Research 2026.

**Chandrasekhar, Arun G, Horacio Larreguy, and Juan Pablo Xandri**, "Testing models of social learning on networks: Evidence from two experiments," *Econometrica*, 2020, *88* (1), 1–32.

**Chaudhry, Zain, Karrar Hussain, and Obeid Ur Rehman**, "Overcoming Intersectarian Divisions through Contact and Leadership: Evidence from a Field Experiment," *Working Paper*, 2025.

**Clochard, Gwen-Jirō**, "Using a brief contact to improve trust in the police by the youth," *Journal of Behavioral and Experimental Economics*, 2025, p. 102376.

**Clochard, Gwen-Jiro, Guillaume Hollard, and Omar Sene**, "Bringing Contact Interventions to the Lab: Effects of Bilateral Discussions on interethnic Trust in Senegal," *Available at SSRN 4693073*, 2023.

**Collier, Paul and Anke Hoeffler**, "Greed and Grievance in Civil War," *Oxford Economic Papers*, 2004, *56* (4), 563–595.

**Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach**, "Not learning from others," Technical Report, National Bureau of Economic Research 2022.

**Corno, Lucia, Eliana La Ferrara, and Justine Burns**, "Interaction, Stereotypes, and Performance: Evidence from South Africa," *American Economic Review*, 2022, *112* (12), 3848–3875.

**Davies, Kristin, Linda R Tropp, Arthur Aron, Thomas F Pettigrew, and Stephen C Wright**, "Cross-group friendships and intergroup attitudes: A meta-analytic review," *Personality and Social Psychology Review*, 2011, *15* (4), 332–351.

**Elwert, Felix, Tamás Keller, and Andreas Kotsadam**, "Rearranging the Desk Chairs: A Large Randomized Field Experiment on the Effects of Close Contact on Interethnic Relations," *American Journal of Sociology*, 2023, *128* (6), 1809–1840.

**Friedman, Willa, Guthrie Gray-Lobe, and Michael Kremer**, "Worker Assignment and National Unity: Are All Stable Matches Socially Stable?," Technical Report, Working Paper 2024.

**Fryer, Roland and Matthew O Jackson**, "A categorical model of cognition and biased decision making," *The BE Journal of Theoretical Economics*, 2008, *8* (1).

**Gee, Laura K, Jason Jones, and Moira Burke**, "Social networks and labor markets: How strong ties relate to job finding on Facebook's social network," *Journal of Labor Economics*, 2017, *35* (2), 485–518.

**Ghosh, Arkadev**, "Religious Divisions and Production Technology: Experimental Evidence from India," *Journal of Political Economy*, 2025.

**_ , Prerna Kundu, Matt Lowe, and Gareth Nellis**, "Creating cohesive communities: A youth camp experiment in india," *Review of Economic Studies*, 2025, p. rdaf026.

**_ , Sam Il Myoung Hwang, and Munir Squires**, "Economic consequences of kinship: Evidence from US bans on cousin marriage," *The Quarterly Journal of Economics*, 2023, *138* (4), 2559–2606.

**Golub, Benjamin and Matthew O Jackson**, "How homophily affects the speed of learning and best-response dynamics," *The Quarterly Journal of Economics*, 2012, *127* (3), 1287–1338.

**Grady, Christopher, Rebecca Wolfe, Danjuma Dawop, and Lisa Inks**, "How Contact Can Promote Societal Change Amid Conflict: An Intergroup Contact Field Experiment in Nigeria," *Proceedings of the National Academy of Sciences*, 2023, *120* (43), e2304882120.

**Granovetter, Mark**, "The Strength of Weak Ties," *American Journal of Sociology*, 1973, *78* (6), 1360–1380.

**Greene, Kenneth F, Erin L Rossiter, Enrique Seira, and Alberto Simpser**, "Interacting as equals reduces partisan polarization in Mexico," *Nature Human Behaviour*, 2025, *9* (1), 147–155.

**Harter, James, Frank Schmidt, and Theodore Hayes**, "Business-Unit-Level Relationship between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis," *Journal of Applied Psychology*, 2002, *87* (2), 268.

**Hewstone, Miles and Rupert Brown**, "Contact is Not Enough: An Intergroup Perspective on the 'Contact Hypothesis'," in Miles Hewstone and Rupert Brown, eds., *Contact and Conflict in Intergroup Encounters*, Oxford: Basil Blackwell, 1986, pp. 1–44.

**Hjort, Jonas**, "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

**Horowitz, Donald**, *The Deadly Ethnic Riot*, Berkeley: University of California Press, 2001.

**Jackson, Matthew O, Stephen M Nei, Erik Snowberg, and Leeat Yariv**, "The dynamics of networks and homophily," Technical Report, National Bureau of Economic Research 2023.

**Johnson, Douglas**, "The Fighting Nuer: Primary Sources and the Origins of a Stereotype," *Africa*, 1981, *51* (1), 508–527.

**Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz**, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

**Lowe, Matt**, "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 2021, *111* (6), 1807–1844.

_ , "Has intergroup contact delivered?," *Annual Review of Economics*, 2025, *17*.

**Macchiavello, Rocco**, "Relational contracts and development," *Annual Review of Economics*, 2022, *14* (1), 337–362.

**Matavelli, Ieda**, "We Don't Talk About Boys: Masculinity Norms Among Adolescents in Brazil," *University of British Columbia, mimeo*, 2024.

**Mobius, Markus and Tanya Rosenblat**, "Social learning in economics," *Annu. Rev. Econ.*, 2014, *6* (1), 827–847.

_ **, Tuan Phan, and Adam Szeidl**, "Treasure hunt: Social learning in the field," Technical Report, National Bureau of Economic Research 2015.

**Mousa, Salma**, "Building Social Cohesion Between Christians and Muslims Through Soccer in Post-ISIS Iraq," *Science*, 2020, *369* (6505), 866–870.

**Okunogbe, Oyebola**, "Does Exposure to Other Ethnic Regions Promote National Integration? Evidence from Nigeria," *American Economic Journal: Applied Economics, forthcoming*, 2023.

**Paluck, Elizabeth and Chelsey Clark**, "Can Playing Together Help Us Live Together?," *Science*, 2020, *369* (6505), 769–770.

**Pettigrew, Thomas**, "Intergroup Contact Theory," *Annual Review of Psychology*, 1998, *49* (1), 65–85.

_ **and Linda Tropp**, "A Meta-Analytic Test of Intergroup Contact Theory," *Journal of Personality and Social Psychology*, 2006, *90* (5), 751.

**Phelps, Edmund S**, "The statistical theory of racism and sexism," *The american economic review*, 1972, *62* (4), 659–661.

**Prasad, Neeraj, Ursula Daxecker, and Kartikeya Batra**, "Sowing Discord: How Violence Transforms Political Identities and Cleavages," *University of Amsterdam, mimeo*, 2024.

**Putnam, Robert**, *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster, 2000.

**Said, Edward**, *Orientalism*, New York: Pantheon Books, 1978.

**Scacco, Alexandra and Shana Warren**, "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria," *American Political Science Review*, 2018, *112* (3), 654–677.

**Schwartzstein, Joshua**, "Selective attention and learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

**Swann, William, Angel Gómez, D Conor Seyle, J Francisco Morales, and Carmen Huici**, "Identity Fusion: the Interplay of Personal and Social Identities in Extreme Group Behavior," *Journal of Personality and Social Psychology*, 2009, *96* (5), 995.

**Tropp, Linda R, Liora Morhayim, and Lusine Grigoryan**, "How Social Categorization Shapes Intergroup Contact: Implications for Understanding Group Members' Subjective Experiences and Prospects for Achieving Attitude Generalization," in "Handbook of Prejudice, Stereotyping, and Discrimination," Routledge, 2024, pp. 476–498.

**Varshney, Ashutosh**, "Ethnic Conflict and Civil Society: India and Beyond," *World Politics*, 2001, *53* (3), 362–398.

**Webb, Duncan**, "Silence to solidarity: Using group dynamics to reduce anti-transgender discrimination in India," 2024.

**Zhou, Yang-Yang and Jason Lyall**, "Prolonged contact does not reshape locals' attitudes toward migrants in wartime settings," *American Journal of Political Science*, 2025, *69* (1), 210–222.

# ONLINE APPENDIX

## "Learning About Outgroups: The Impact of Broad Versus Deep Interactions"

Anujit Chakraborty[&] Arkadev Ghosh[#] Matt Lowe[+] Gareth Nellis[*]

[&]*University of California, Davis* [#]*Duke University*
[+]*University of British Columbia* [*]*University of Pennsylvania*

## A Supplementary Tables

Table S1: Characterizing sample selection across recruitment and screening waves.

| | Sample mean (SD); Completed... | | | Regression estimates: | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | (2) vs. (1) | | (3) vs. (2) | | (3) vs. (1) | |
| Variable: | Baseline 1 (1) | Baseline 2 (2) | Lead-in (3) | Δ (4) | *p*-value (5) | Δ (6) | *p*-value (7) | Δ (8) | *p*-value (9) |
| (a) Age | 22.14 | 21.81 | 21.65 | -0.33 | 0.08 | -0.16 | 0.42 | -0.49 | 0.01 |
| | (5.12) | (4.73) | (4.63) | | | | | | |
| (b) Completed primary school | 0.77 | 0.81 | 0.86 | 0.04 | 0.01 | 0.05 | <0.01 | 0.09 | <0.01 |
| | (0.42) | (0.39) | (0.35) | | | | | | |
| (c) Household income | 12890.47 | 12768.21 | 12870.85 | -122.25 | 0.62 | 102.64 | 0.71 | -19.62 | 0.94 |
| | (6518.72) | (6422.16) | (6367.81) | | | | | | |
| (d) Currently employed | 0.30 | 0.27 | 0.23 | -0.03 | 0.09 | -0.03 | 0.07 | -0.06 | <0.01 |
| | (0.46) | (0.44) | (0.42) | | | | | | |
| (e) Married | 0.16 | 0.14 | 0.12 | -0.02 | 0.25 | -0.02 | 0.24 | -0.03 | 0.02 |
| | (0.36) | (0.35) | (0.33) | | | | | | |
| (f) Number of people in household | 4.29 | 4.20 | 4.14 | -0.09 | 0.18 | -0.06 | 0.32 | -0.14 | 0.03 |
| | (1.90) | (1.42) | (1.33) | | | | | | |
| (g) Number of children desired | 1.56 | 1.53 | 1.53 | -0.02 | 0.38 | 0.00 | 0.85 | -0.03 | 0.31 |
| | (0.69) | (0.59) | (0.58) | | | | | | |
| (h) Hours worked each week | 12.27 | 11.24 | 9.66 | -1.03 | 0.16 | -1.58 | 0.04 | -2.61 | <0.01 |
| | (19.15) | (18.65) | (17.66) | | | | | | |
| (i) Trust rank: Hindus | 1.65 | 1.61 | 1.61 | -0.04 | 0.34 | -0.01 | 0.87 | -0.04 | 0.28 |
| | (0.99) | (0.96) | (0.94) | | | | | | |
| (j) Trust rank: Muslims | 2.46 | 2.50 | 2.51 | 0.03 | 0.55 | 0.02 | 0.80 | 0.05 | 0.42 |
| | (1.52) | (1.51) | (1.53) | | | | | | |
| (k) Feeling Thermometer: Narendra Modi | 66.03 | 67.08 | 67.46 | 1.06 | 0.41 | 0.37 | 0.79 | 1.43 | 0.30 |
| | (33.43) | (33.16) | (33.21) | | | | | | |
| Sample *N* | 1,479 | 1,254 | 964 | | | | | | |

*Notes:* This table shows the characteristics of the sample across different recruitment and screening stages. Columns 1 to 3 present simple averages of variables measured for all subjects during baseline 1 (with standard deviations in parentheses). Monthly household income in INR in row (c) is a continuous variable coded as the midpoint of a selected income bin, or the minimum value in the case of the highest bin (>50,000 INR). Trust rankings in rows (i) and (j) are measured from 1 (most trust) to 5 (least trust). Participants ranked five groups: Hindus, Muslims, Sikhs, people from Nepal, and people from North Indian states. The feeling thermometer in row (k) ranges from 0 (least warm) to 100 (most warm). Columns 4 to 9 estimate the average differences between pairs of samples. Differences and *p*-values come from bivariate regressions of the baseline 1 variable on an indicator for whether subjects completed the later of the two screening steps. We use robust standard errors.

Table S2: Baseline summary statistics for the experimental sample.

| Variable: | N | Mean | St. Dev. | Min. | Pctl. 25 | Median | Pctl. 75 | Max. |
|---|---|---|---|---|---|---|---|---|
| (a) Age | 918 | 22 | 4.7 | 18 | 18 | 20 | 23 | 40 |
| (b) Completed primary school | 918 | 0.86 | 0.35 | 0 | 1 | 1 | 1 | 1 |
| (c) Completed 12th standard | 918 | 0.65 | 0.48 | 0 | 0 | 1 | 1 | 1 |
| (d) Completed university | 918 | 0.12 | 0.33 | 0 | 0 | 0 | 0 | 1 |
| (e) Household income | 918 | 12900 | 6315 | 2500 | 7500 | 12500 | 17500 | 50000 |
| (f) Currently employed | 918 | 0.23 | 0.42 | 0 | 0 | 0 | 0 | 1 |
| (g) Has bank account | 918 | 0.87 | 0.33 | 0 | 1 | 1 | 1 | 1 |
| (h) Has smartphone | 918 | 0.93 | 0.25 | 0 | 1 | 1 | 1 | 1 |
| (i) Has Facebook | 918 | 0.84 | 0.36 | 0 | 1 | 1 | 1 | 1 |
| (j) Married | 918 | 0.13 | 0.33 | 0 | 0 | 0 | 0 | 1 |
| (k) Number of people in household | 906 | 4.2 | 1.3 | 1 | 3 | 4 | 5 | 18 |
| (l) Number of children desired | 917 | 1.5 | 0.57 | 0 | 1 | 2 | 2 | 5 |
| (m) Hours worked each week | 918 | 9.6 | 18 | 0 | 0 | 0 | 12 | 72 |
| (n) Trust rank: Hindus | 918 | 1.6 | 0.93 | 1 | 1 | 1 | 2 | 5 |
| (o) Trust rank: Muslims | 918 | 2.5 | 1.5 | 1 | 1 | 2 | 4 | 5 |
| (p) Ward exposure: Hindus | 918 | 61 | 28 | 0 | 40 | 66 | 86 | 100 |
| (q) Ward exposure: Muslims | 918 | 37 | 28 | 0 | 10 | 30 | 60 | 100 |
| (r) Feeling Thermometer: Narendra Modi | 918 | 67 | 33 | 0 | 48 | 80 | 100 | 100 |
| (s) Interest in attending social event | 918 | 1.7 | 0.48 | 0 | 1 | 2 | 2 | 2 |
| (t) Enjoys meeting new people | 918 | 2.6 | 0.63 | 0 | 2 | 3 | 3 | 3 |
| (u) Depression score | 918 | 2.4 | 2.6 | 0 | 0 | 2 | 4 | 16 |
| (v) Hindu | 918 | 0.62 | 0.49 | 0 | 0 | 1 | 1 | 1 |

*Notes:* This table presents summary statistics for variables collected in baseline surveys 1 and 2. Only subjects who completed the endline survey are included. Monthly household income in INR in row (e) is a continuous variable coded as the midpoint of a selected income bin, or the minimum value in the case of the highest bin (>50,000 INR). Trust rankings in rows (n) and (o) are measured from 1 (most trust) to 5 (least trust). Participants ranked five groups: Hindus, Muslims, Sikhs, people from Nepal, and people from North Indian states. Ward exposure in rows (p) and (q) is the number of people from that group, out of 100, that the subject estimates to be living in his ward. The feeling thermometer in row (r) ranges from 0 (least warm) to 100 (most warm). Interest in row (s) ranges from 0 (not at all interested) to 2 (very interested). The measure in row (t) is how much subjects agree with the statement that they enjoy meeting new people; it ranges from 0 (very much disagree) to 3 (very much agree). The depression score in row (u) is the sum of answers to the eight PHQ-8 questions. One example question is "over the last two weeks, how often have you been feeling down, depressed, or hopeless?" Answers range from 0 (not at all) to 3 (nearly every day). The total score then ranges from 0 to 24.

Table S3: Randomization checks.

| | Control | | Ingroup deep vs. control | | Outgroup deep vs. control | | Outgroup broad vs. control | | (3) = (5) | (3) = (7) | (5) = (7) | (3) = (5) = (7) = 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean (1) | SD (2) | $\beta$ (3) | $p$-value (4) | $\beta$ (5) | $p$-value (6) | $\beta$ (7) | $p$-value (8) | $p$-value (9) | $p$-value (10) | $p$-value (11) | $p$-value (12) |
| (a) Age | 21.57 | 4.70 | 0.72 | 0.14 | -0.04 | 0.91 | -0.02 | 0.96 | 0.10 | 0.12 | 0.95 | 0.35 |
| (b) Completed primary school | 0.88 | 0.32 | -0.05 | 0.10 | -0.05 | 0.12 | -0.02 | 0.49 | 0.93 | 0.34 | 0.35 | 0.29 |
| (c) Completed 12th standard | 0.63 | 0.48 | -0.02 | 0.69 | 0.06 | 0.18 | 0.03 | 0.42 | 0.10 | 0.26 | 0.59 | 0.34 |
| (d) Completed university | 0.10 | 0.31 | 0.01 | 0.82 | 0.04 | 0.18 | 0.01 | 0.69 | 0.37 | 0.90 | 0.34 | 0.59 |
| (e) Household income | 12262.44 | 5618.06 | 1441.72 | 0.02 | 616.82 | 0.26 | 418.30 | 0.45 | 0.20 | 0.13 | 0.73 | 0.16 |
| (f) Currently employed | 0.24 | 0.43 | 0.03 | 0.49 | 0.01 | 0.78 | -0.02 | 0.63 | 0.64 | 0.24 | 0.42 | 0.69 |
| (g) Has bank account | 0.83 | 0.37 | 0.07 | 0.05 | 0.03 | 0.33 | 0.05 | 0.09 | 0.22 | 0.57 | 0.45 | 0.17 |
| (h) Has smartphone | 0.94 | 0.24 | -0.01 | 0.79 | 0.00 | 0.90 | -0.01 | 0.52 | 0.86 | 0.82 | 0.60 | 0.93 |
| (i) Has Facebook | 0.84 | 0.37 | 0.03 | 0.45 | 0.00 | 0.92 | 0.00 | 0.91 | 0.41 | 0.50 | 0.82 | 0.86 |
| (j) Married | 0.12 | 0.33 | 0.04 | 0.18 | -0.01 | 0.67 | 0.01 | 0.62 | 0.08 | 0.38 | 0.34 | 0.33 |
| (k) Number of people in household | 4.19 | 1.41 | 0.03 | 0.84 | -0.08 | 0.53 | 0.05 | 0.70 | 0.40 | 0.86 | 0.31 | 0.74 |
| (l) Number of children desired | 1.53 | 0.55 | 0.02 | 0.73 | 0.03 | 0.50 | 0.00 | 0.97 | 0.84 | 0.71 | 0.48 | 0.88 |
| (m) Hours worked each week | 9.48 | 17.81 | 3.09 | 0.08 | 0.35 | 0.82 | -0.96 | 0.53 | 0.09 | 0.01 | 0.35 | 0.09 |
| (n) Trust rank: Hindus | 1.70 | 0.95 | -0.03 | 0.54 | -0.02 | 0.76 | 0.03 | 0.64 | 0.82 | 0.25 | 0.44 | 0.69 |
| (o) Trust rank: Muslims | 2.34 | 1.48 | -0.03 | 0.80 | 0.10 | 0.26 | -0.07 | 0.43 | 0.30 | 0.77 | 0.05 | 0.25 |
| (p) Ward exposure: Hindus | 60.93 | 27.45 | -2.58 | 0.17 | -3.11 | 0.09 | -1.91 | 0.30 | 0.77 | 0.71 | 0.48 | 0.36 |
| (q) Ward exposure: Muslims | 37.12 | 27.19 | 2.49 | 0.19 | 3.81 | 0.03 | 2.45 | 0.19 | 0.47 | 0.98 | 0.43 | 0.21 |
| (r) Feeling Thermometer: Narendra Modi | 63.99 | 35.80 | 1.95 | 0.40 | 0.49 | 0.83 | -1.48 | 0.54 | 0.51 | 0.14 | 0.39 | 0.53 |
| (s) Interest in attending social event | 1.64 | 0.50 | 0.01 | 0.87 | 0.03 | 0.58 | 0.02 | 0.70 | 0.73 | 0.84 | 0.86 | 0.95 |
| (t) Enjoys meeting new people | 2.50 | 0.64 | 0.06 | 0.38 | 0.04 | 0.49 | 0.07 | 0.21 | 0.80 | 0.85 | 0.60 | 0.63 |
| (u) Depression score | 2.26 | 2.59 | 0.08 | 0.76 | 0.50 | 0.05 | 0.17 | 0.47 | 0.13 | 0.71 | 0.17 | 0.24 |

*Notes:* This table tests whether baseline covariates are balanced across experimental conditions. Each row presents the results of a regression of one baseline covariate on the three treatment indicators, and strata fixed effects. We use robust standard errors. Columns 3–8 report the coefficients and $p$-values from those regressions. Columns 9–11 report $p$-values testing the equality of those coefficients to one another. Column 12 presents the $p$-value from a joint test of the statistical significance of all treatment indicators. Monthly household income in INR in row (e) is a continuous variable coded as the midpoint of an income bin or the minimum value for the highest bin. Trust rankings in rows (n) and (o) are measured from 1 (most trust) to 5 (least trust). Participants ranked five groups: Hindus, Muslims, Sikhs, people from Nepal, and people from North Indian states. Ward exposure in rows (p) and (q) are the number of people, out of 100, the subject reports as living in their ward. The feeling thermometer in row (r) ranges from 0 (least warm) to 100 (most warm). Interest in row (s) ranges from 0 (not at all interested) to 2 (very interested) and is treated as a continuous variable. The measure in row (t) is how much subjects agree with the statement that they enjoy meeting new people; it ranges from 0 (Very much disagree) to 3 (Very much agree) and is treated as a continuous variable. The depression score in row (u) is the sum of answers to the eight PHQ-8 questions. One example question is "over the last two weeks, how often have you been feeling down, depressed, or hopeless?" Answers range from 0 (not at all) to 3 (nearly every day). The total score then ranges from 0 to 24.

Table S4: Attrition and attendance checks.

| | Outcome: | | | |
|---|---|---|---|---|
| | Attrited from endline | | Attended work | |
| | (1) | (2) | (3) | (4) |
| Ingroup, deep | -0.02 | | | |
| | (0.02) | | | |
| Outgroup, deep | -0.01 | | 0.00 | |
| | (0.02) | | (0.02) | |
| Outgroup, broad | -0.03 | | 0.02 | |
| | (0.02) | | (0.02) | |
| Hindu | | 0.00 | | -0.01 |
| | | (0.01) | | (0.01) |
| | | | | |
| *N* | 964 | 964 | 4,368 | 4,368 |
| Sample | Full | Full | Treated only | Treated only |
| Randomization strata FEs | Yes | No | Yes | No |
| Control mean | 0.06 | | | |
| Muslim mean | | 0.05 | | 0.96 |
| Ingroup, deep mean | | | 0.94 | |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.41 | | 0.06 | |

*Notes:* This table tests for differential attendance and attrition across experimental treatment arms. The unit of analysis in columns 1 and 2 is the subject and the sample includes all subjects who were randomized. The outcome in columns 1 and 2 is a dummy variable that takes one if the subject did not complete the endline survey. The unit of analysis in columns 3 and 4 is the worker/day (six work days per worker in total), and the sample includes all subjects who were randomized to work. The outcome in columns 3 and 4 is a dummy variable that takes one if the subject attended work on a given day. Standard errors are shown in parentheses; these are robust in columns 1 and 2, and are clustered by subject in columns 3 and 4. *p<0.1; **p<0.05; ***p<0.01.

Table S5: Purpose of study checks.

| | Outcome: Mentioned... | |
|---|---|---|
| | Hindu-Muslim relations (1) | Unity/ community (2) |
| Ingroup, deep | -0.02 | -0.01 |
| | (0.03) | (0.03) |
| Outgroup, deep | 0.01 | 0.01 |
| | (0.03) | (0.02) |
| Outgroup, broad | 0.04 | -0.01 |
| | (0.03) | (0.02) |
| | | |
| *N* | 918 | 918 |
| Control, mean | 0.10 | 0.07 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.34 | 0.44 |

*Notes:* This table tests whether endline subjects' reports of the purpose of the study differ according to experimental treatment arm. Subjects were asked at endline, "What do you think was the purpose of this research study? What was it trying to understand?" Subjects' qualitative responses were recorded by enumerators and categorized using ChatGPT5. The outcome in column 1 is a binary variable for whether the subject gave Hindu-Muslim relations as the reason for the study. The outcome in column 2 is a binary variable for whether the subject gave Unity/community as the reason for the study. The unit of analysis is the subject. Regressions include randomization strata fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S6: Subjects' mood is worse in *Outgroup, broad*, even on the first day of the intervention.

| | Outcome ($z$-scores): | | |
|---|---|---|---|
| | Happiness today (1) | Partner relationship (2) | Partner close friend (3) |
| Outgroup, deep | 0.24** | 0.13 | 0.15 |
| | (0.11) | (0.12) | (0.12) |
| Outgroup, broad | 0.09 | 0.17 | -0.02 |
| | (0.11) | (0.12) | (0.12) |
| | | | |
| *N* | 673 | 674 | 674 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.09 | 0.73 | 0.05 |

*Notes:* This table reports estimates of treatment effects on subjects' mood and partner relationship on the first day of the intervention. The survey questions for each column are: (1) Which picture best describes your emotions during today's work? (0 = Very sad face to 4 = Very happy face); (2) Which picture best describes your relationship with your partner today? (0 = Self and partner circles apart to 4 = Self and partner circles fully overlapping); (3) How close of a friend do you consider your partner to be? (1 = Not at all to 10 = To a great extent). All outcomes are $z$-scores, centered and standardized according to the mean and standard deviation of the variable in the *Ingroup, deep* condition. Only subjects assigned to work are included in the regressions. We drop surveys for subjects whose partners were absent that day. The unit of analysis is the subject. Regressions include religion × wave × day × shift fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S7: Effects on endline recall of workplace happiness, productivity, and partners.

| | | Team bags, last day | | Partners' names | |
|---|---|---|---|---|---|
| | Overall happiness (1) | Absolute error (2) | Within 5 bags (3) | Share claimed (4) | Share correct (5) |
| Outgroup, deep | 0.00 | 0.47 | -0.00 | -0.07** | -0.13*** |
| | (0.10) | (0.88) | (0.05) | (0.03) | (0.04) |
| Outgroup, broad | -0.07 | 0.23 | 0.00 | -0.35*** | -0.49*** |
| | (0.10) | (0.82) | (0.05) | (0.03) | (0.04) |
| | | | | | |
| *N* | 697 | 647 | 647 | 697 | 697 |
| Ingroup, deep mean | 3.65 | 5.00 | 0.64 | 0.90 | 0.82 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.46 | 0.70 | 0.98 | <0.01 | <0.01 |

*Notes:* This table shows effects on recall at endline, using the questions (1) "Think back to the final 6-day work period. In general, how happy or unhappy did you feel during those 6 days?" (0–4). (2)–(3) Recalled number of bags made with the partner on the last day (big and small bags each count as one), compared with the booth's actual count: absolute error, and whether the guess is within 5 bags (incentivized); missing if either partner was absent. (4) Share of partners whose name the subject says they remember (deep: answered yes; broad: names given out of six). (5) Share of partners named correctly, matching first names and allowing for spelling and transliteration differences. Regressions include randomization strata fixed effects; *Ingroup, deep* is the omitted group. Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S8: Heterogeneous effects of *Outgroup, broad* on ties with former work partners, by religion.

| | Outcome: | | |
|---|---|---|---|
| | Had contact with any work partner within past week (1) | Received help finding paid work from any partner (2) | Gave financial assistance to any partner (3) |
| **Panel A: Hindus** | | | |
| Outgroup, broad | -0.16*** | 0.01 | -0.04** |
| | (0.06) | (0.04) | (0.02) |
| *N* | 253 | 253 | 253 |
| Outgroup, deep, outcome mean | 0.65 | 0.14 | 0.04 |
| **Panel B: Muslims** | | | |
| Outgroup, broad | -0.11* | -0.09** | -0.01 |
| | (0.06) | (0.04) | (0.01) |
| *N* | 252 | 252 | 252 |
| Outgroup, deep, outcome mean | 0.54 | 0.14 | 0.02 |

*Notes:* This table shows estimated treatment effects on recent contact with former work partners, by the subject's religion. The sample includes subjects assigned to *Outgroup, deep* or *Outgroup, broad* only. The outcomes are: (a) A binary variable for whether the subject reports having had contact (in-person or by phone) with (any of) their work partner(s) within the last week; (b) Whether the subject reports having received help finding paid work from (any of) their work partner(s); (c) Whether the subject reports having borrowed from, or lent to, (any of) their work partner(s). The unit of analysis is the subject. Regressions include randomization strata fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S9: *Outgroup, broad* increases prediction accuracy about the more visible preferences and behaviors of outgroup strangers: tabulated results and robustness.

| | Full sample | | | Hindus | | | Muslims | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Index components:* | | | *Index components:* | | | *Index components:* | | |
| | All | High visibility | Low visibility | All | High visibility | Low visibility | All | High visibility | Low visibility |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Panel A: Indexes with only binary components** | | | | | | | | | |
| Outgroup, broad | 0.04 | 0.10** | -0.04 | 0.09** | 0.21*** | -0.04 | -0.01 | 0.00 | -0.03 |
| | (0.03) | (0.04) | (0.05) | (0.04) | (0.06) | (0.07) | (0.04) | (0.06) | (0.07) |
| *N* | 505 | 505 | 505 | 253 | 253 | 253 | 252 | 252 | 252 |
| **Panel B: Indexes with binary and continuous components** | | | | | | | | | |
| Outgroup, broad | 0.04 | 0.09** | -0.02 | 0.09** | 0.18*** | -0.02 | -0.01 | -0.01 | -0.02 |
| | (0.03) | (0.04) | (0.05) | (0.04) | (0.06) | (0.06) | (0.04) | (0.05) | (0.07) |
| *N* | 505 | 505 | 505 | 253 | 253 | 253 | 252 | 252 | 252 |

*Notes:* This table shows estimated treatment effects on the accuracy of predictions made about outgroups in the holdout sample. The unit of analysis is the subject. The sample is restricted to subjects assigned to the *Outgroup, deep* and *Outgroup, broad* conditions. All outcomes are *z*-score indexes, averaging components that are centered and standardized according to the variable's mean and standard deviation in *Outgroup, deep* condition. The outcome for columns 1, 4, and 7 is the mean of the nine components measuring prediction accuracy, as described in the main text. The outcome for columns 2, 5, and 8 is the mean of the five components: production, identity, bag, children, and sanitizer. The outcome for columns 3, 6, and 9 is the mean of the remaining four components: movie, risk, dictator, and coin. In panel A, all components (prior to standardization) are binary indicators for prediction accuracy. In panel B, the raw production, children, dictator, and coin components are treated as continuous; we take the absolute difference between the guess and "truth" benchmark, we add the integer one, we take the log of the resulting value (so that closer guesses receive more weight), and we then center and standardize the variable according to mean and standard deviation of the variable in the *Outgroup, deep* condition. Regressions include randomization strata fixed effects. The 95% confidence intervals and *p*-values for the differences are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

Table S10: No effect of contact on political outcomes.

| | Outcome (*z*-scores): | | |
|---|---|---|---|
| | National identity (1) | Democracy rank (2) | Modi thermometer (3) |
| Ingroup, deep | 0.01 | 0.13 | 0.06 |
| | (0.10) | (0.11) | (0.07) |
| Outgroup, deep | 0.03 | 0.15 | 0.07 |
| | (0.09) | (0.10) | (0.06) |
| Outgroup, broad | -0.02 | 0.16* | 0.08 |
| | (0.09) | (0.09) | (0.07) |
| *N* | 918 | 918 | 918 |
| Control, raw outcome mean | 3.70 | 1.71 | 63.24 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.58 | 0.87 | 0.90 |

*Notes:* This table presents treatment effect estimates on political attitudes. The unit of analysis is the subject. The outcomes for each column are based on the following endline survey questions: (1) Let us suppose that you had to choose between being an Indian and being a [Hindu/Muslim]. Which of these two identities do you feel most strongly attached to? (1 = Only Indian, 5 = Only [Hindu/Muslim]); (2) I'm going to describe various types of political systems. Please rank how good you think the systems are as a way of governing this country. Rank 1 means "best," rank 2 means "second best," etc.: Having a democratic political system; (3) I'd like to get your feelings toward some famous Indians. I'll read the name and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favourable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person: Narendra Modi. All outcomes are *z*-scores, centered and standardized according to the mean and standard deviation of the variable in the control group. Regressions include randomization strata fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S11: Updating about partners.

| | More visible | | | | Less visible | | | |
|---|---|---|---|---|---|---|---|---|
| | Kids guess (1) | Kids accuracy (2) | Identity guess (3) | Identity accuracy (4) | Dictator guess (5) | Dictator accuracy (6) | Coin guess (7) | Accuracy index (8) |
| **A. Hindus** | | | | | | | | |
| Outgroup, deep | -0.24** | 0.10** | 0.02 | 0.04 | 0.62 | 0.03 | 0.04 | 0.19** |
| | (0.10) | (0.04) | (0.10) | (0.06) | (2.41) | (0.05) | (0.21) | (0.08) |
| Outgroup, broad | -0.32*** | 0.05 | 0.10 | 0.02 | -2.58 | -0.01 | 0.21 | 0.07 |
| | (0.10) | (0.04) | (0.10) | (0.06) | (2.48) | (0.05) | (0.22) | (0.09) |
| *N* | 373 | 373 | 373 | 373 | 370 | 370 | 373 | 373 |
| Control, outcome mean | 2.17 | 0.86 | 1.99 | 0.32 | 31.20 | 0.18 | 6.16 | -0.17 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.30 | 0.15 | 0.45 | 0.67 | 0.18 | 0.44 | 0.45 | 0.14 |
| **B. Muslims** | | | | | | | | |
| Outgroup, deep | 0.05 | -0.04* | -0.19** | 0.01 | 4.82** | -0.07 | 0.14 | -0.10 |
| | (0.07) | (0.02) | (0.09) | (0.07) | (2.38) | (0.06) | (0.21) | (0.07) |
| Outgroup, broad | 0.15** | -0.03 | -0.11 | 0.07 | 2.92 | -0.07 | 0.20 | -0.05 |
| | (0.07) | (0.02) | (0.09) | (0.07) | (2.21) | (0.06) | (0.22) | (0.07) |
| *N* | 353 | 353 | 353 | 353 | 347 | 347 | 353 | 353 |
| Control, outcome mean | 1.62 | 0.99 | 2.41 | 0.39 | 24.82 | 0.25 | 5.92 | 0.12 |
| Outgroup, deep vs. Outgroup, broad *p*-value | 0.15 | 0.74 | 0.36 | 0.37 | 0.43 | 0.97 | 0.77 | 0.44 |

*Notes:* $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors in parentheses. This table presents treatment effects on participants' endline guesses about their partners' characteristics. We do not observe partners' true outcomes in the dishonesty (coin) game, so accuracy of partner predictions cannot be computed for this measure. Nevertheless, shifts in predicted values may indicate updating about the partner. The accuracy index (in the final column) is therefore constructed from predictions about partners' (i) desired number of children ("Kids"), (ii) national versus religious self-identification ("Identity"), and (iii) dictator-game giving to an outgroup stranger ("Dictator"). Results are reported separately for Hindu and Muslim participants. The omitted category is the control group, whose members made predictions about randomly matched outgroup treated participants. All regressions include randomization strata fixed effects.

Table S12: Beliefs about strangers and partners' productivity: robustness.

| | Main (1) | Hindus (2) | Muslims (3) | Day-3 signal (4) | Realized partners (5) |
|---|---|---|---|---|---|
| **Panel A: Outgroup minus ingroup guess** | | | | | |
| Outgroup, deep slope | -0.03 | 0.12 | -0.15 | -0.00 | -0.04 |
| | (0.10) | (0.09) | (0.15) | (0.08) | (0.10) |
| Outgroup, broad slope | 0.74*** | 0.84** | 0.58* | 0.50*** | 0.63*** |
| | (0.26) | (0.39) | (0.33) | (0.19) | (0.23) |
| Broad − deep | 0.77*** | 0.72* | 0.73* | 0.50** | 0.67*** |
| | (0.28) | (0.41) | (0.37) | (0.21) | (0.25) |
| **Panel B: Guess about an outgroup stranger** | | | | | |
| Outgroup, deep slope | 0.15 | 0.17 | 0.13 | 0.16** | 0.15 |
| | (0.10) | (0.15) | (0.14) | (0.07) | (0.10) |
| Outgroup, broad slope | 0.69** | 0.72 | 0.59 | 0.45* | 0.52* |
| | (0.34) | (0.64) | (0.38) | (0.24) | (0.30) |
| Broad − deep | 0.54 | 0.55 | 0.45 | 0.29 | 0.38 |
| | (0.36) | (0.69) | (0.41) | (0.26) | (0.32) |
| **Panel C: Guess about an ingroup stranger** | | | | | |
| Outgroup, deep slope | 0.18 | 0.05 | 0.28 | 0.16 | 0.19 |
| | (0.14) | (0.16) | (0.21) | (0.11) | (0.14) |
| Outgroup, broad slope | -0.06 | -0.12 | 0.01 | -0.05 | -0.10 |
| | (0.35) | (0.68) | (0.39) | (0.25) | (0.30) |
| Broad − deep | -0.23 | -0.17 | -0.27 | -0.21 | -0.29 |
| | (0.38) | (0.73) | (0.44) | (0.28) | (0.34) |
| *N* | 505 | 253 | 252 | 505 | 497 |

*Notes:* Each column reports slopes of endline productivity guesses (bags per hour) on the lead-in productivity of subjects' assigned partners (the single partner in Outgroup, deep; the average of the six partners in Outgroup, broad), measured relative to the average lead-in productivity of all subjects of the partners' religion assigned to work in the same wave and shift. Column 1 is the main specification (Figure 7); columns 2 and 3 split the sample by subject religion; column 4 measures productivity on the third lead-in day only; column 5 uses the partners subjects actually worked with rather than those assigned. All regressions control for the pool average, the subject's own lead-in productivity, and randomization strata fixed effects. Balance: partners' relative productivity is unrelated to subjects' own lead-in productivity and to their baseline ingroup bias in productivity beliefs (all $p > 0.3$). Heteroskedasticity-robust standard errors in parentheses.

Table S13: Summary statistics for demographic characteristics of the holdout sample.

| Variable: | N | Mean | St. Dev. | Min. | Pctl. 25 | Median | Pctl. 75 | Max. |
|---|---|---|---|---|---|---|---|---|
| (a) Age | 165 | 26 | 6.6 | 18 | 20 | 24 | 30 | 41 |
| (b) Completed primary school | 165 | 0.59 | 0.49 | 0 | 0 | 1 | 1 | 1 |
| (c) Household income | 165 | 15545 | 8111 | 2500 | 12500 | 12500 | 17500 | 50000 |
| (d) Married | 165 | 0.38 | 0.49 | 0 | 0 | 0 | 1 | 1 |
| (e) Number of people in household | 165 | 4.6 | 2.1 | 2 | 3 | 4 | 5 | 19 |
| (f) Currently employed | 165 | 0.74 | 0.44 | 0 | 0 | 1 | 1 | 1 |
| (g) Hours worked each week | 165 | 36 | 26 | 0 | 0 | 42 | 56 | 84 |
| (h) Hindu | 165 | 0.59 | 0.49 | 0 | 0 | 1 | 1 | 1 |

*Notes:* This table presents summary statistics for the holdout sample. Only subjects who were guessed about are included in the sample. Monthly household income in INR in row (c) is a continuous variable coded as the midpoint of a selected income bin, or the minimum value in the case of the highest bin (>50,000 INR).

Table S14: *Outgroup, broad* versus *Outgroup, deep*: adjusting for multiple hypothesis testing.

| Outcome | Source | $p$-value | $p$-value, network-clustered | Sharpened $q$-value (within domain) | Sharpened $q$-value (all tests) |
|---|---|---|---|---|---|
| **Panel A: Interactions during work** | | | | | |
| Happiness | Table 3 (1) | 0.062 | 0.075 | 0.047 | 0.131 |
| Partner relationship | Table 3 (2) | <0.001 | <0.001 | 0.001 | 0.001 |
| Partner close friend | Table 3 (3) | <0.001 | <0.001 | 0.001 | 0.001 |
| Own performance | Table 3 (4) | <0.001 | <0.001 | 0.001 | 0.001 |
| Partner's performance | Table 3 (5) | <0.001 | <0.001 | 0.001 | 0.001 |
| Happiness, day 1 | Table S6 (1) | 0.091 | 0.094 | 0.057 | 0.177 |
| Partner relationship, day 1 | Table S6 (2) | 0.726 | 0.719 | 0.115 | 0.823 |
| Partner close friend, day 1 | Table S6 (3) | 0.049 | 0.062 | 0.043 | 0.115 |
| Number of conversation topics | Table 4 (7) | 0.030 | 0.051 | 0.031 | 0.086 |
| **Panel B: Productivity** | | | | | |
| Number of small bags | Table 7 (1) | 0.653 | 0.685 | 1.000 | 0.823 |
| Small bag quality | Table 7 (2) | 0.828 | 0.843 | 1.000 | 0.883 |
| Small bags × quality | Table 7 (3) | 0.752 | 0.775 | 1.000 | 0.823 |
| Number of big bags | Table 7 (4) | 0.630 | 0.664 | 1.000 | 0.823 |
| Big bag quality | Table 7 (5) | 0.711 | 0.735 | 1.000 | 0.823 |
| Big bags × quality | Table 7 (6) | 0.866 | 0.878 | 1.000 | 0.895 |
| **Panel C: Ties with partners after work** | | | | | |
| Contact with any partner, past week | Figure 3 (a) | 0.002 | 0.007 | 0.006 | 0.010 |
| Help finding paid work from any partner | Figure 3 (b) | 0.192 | 0.256 | 0.069 | 0.317 |
| Borrowed from or lent to any partner | Figure 3 (c) | 0.028 | 0.071 | 0.030 | 0.086 |
| **Panel D: Beliefs about outgroup strangers** | | | | | |
| Accuracy, all components: Full sample | Figure 5 (a)(i) | 0.195 | 0.221 | 0.306 | 0.317 |
| Accuracy, high visibility: Full sample | Figure 5 (b)(i) | 0.013 | 0.010 | 0.038 | 0.050 |
| Accuracy, low visibility: Full sample | Figure 5 (c)(i) | 0.449 | 0.509 | 0.639 | 0.674 |
| Accuracy, all components: Hindus | Figure 5 (a)(ii) | 0.032 | 0.046 | 0.060 | 0.086 |
| Accuracy, high visibility: Hindus | Figure 5 (b)(ii) | <0.001 | <0.001 | 0.007 | 0.005 |
| Accuracy, low visibility: Hindus | Figure 5 (c)(ii) | 0.509 | 0.545 | 0.639 | 0.715 |
| Accuracy, all components: Muslims | Figure 5 (a)(iii) | 0.752 | 0.768 | 0.719 | 0.823 |
| Accuracy, high visibility: Muslims | Figure 5 (b)(iii) | 0.960 | 0.962 | 0.924 | 0.976 |
| Accuracy, low visibility: Muslims | Figure 5 (c)(iii) | 0.681 | 0.674 | 0.719 | 0.823 |
| Response to partners' productivity | Figure 7 (a) | 0.006 | 0.017 | 0.026 | 0.024 |
| **Panel E: Beliefs about partners** | | | | | |
| Productivity prediction accuracy | Table 6 (2) | 0.894 | 0.896 | 1.000 | 0.899 |
| Accuracy index, Hindus | Table S11 (A) | 0.136 | 0.135 | 0.689 | 0.250 |
| Accuracy index, Muslims | Table S11 (B) | 0.443 | 0.498 | 0.795 | 0.674 |
| **Panel F: Behavior toward outgroup strangers** | | | | | |
| Outgroup friends | Table 8 (1) | 0.598 | 0.583 | 1.000 | 0.823 |
| Reservation wage | Table 8 (2) | 0.470 | 0.412 | 1.000 | 0.674 |
| Dictator giving | Table 8 (3) | 0.468 | 0.478 | 1.000 | 0.674 |
| Replied WhatsApp | Table 8 (4) | 0.833 | 0.841 | 1.000 | 0.883 |

*Notes:* This table explores the robustness of the *Outgroup, broad* versus *Outgroup, deep* comparisons summarized in Table 1 to standard error clustering and multiple hypothesis testing. Each row is one test of the difference between *Outgroup, broad* and *Outgroup, deep*, estimated as in the table or figure (and column or panel) given in the second column. The third column reports the unadjusted $p$-value. The fourth column reports the $p$-value with standard errors clustered by partner network (the deep pair, the broad rotation batch, or the individual for the control group). The last two columns report sharpened $q$-values, which control the false discovery rate using the two-step procedure of Benjamini et al. (2006), implemented as in Anderson (2008). A $q$-value is the smallest false discovery rate at which the null hypothesis of no difference is rejected. "Within domain" treats each panel, which corresponds to an outcome domain in Table 1, as a separate family of hypotheses; "all tests" treats all 35 tests as a single family. $q$-values are computed in steps of 0.001, so 0.001 is the smallest attainable value. Note that because the procedure adapts to the number of hypotheses rejected within a family, the raw $p$-value can be larger than the sharpened $q$-value when other hypotheses in the family are strongly rejected.

## B Supplementary Figures

Figure S1: Photographs of the big and small bags that formed the production task for the experiment.

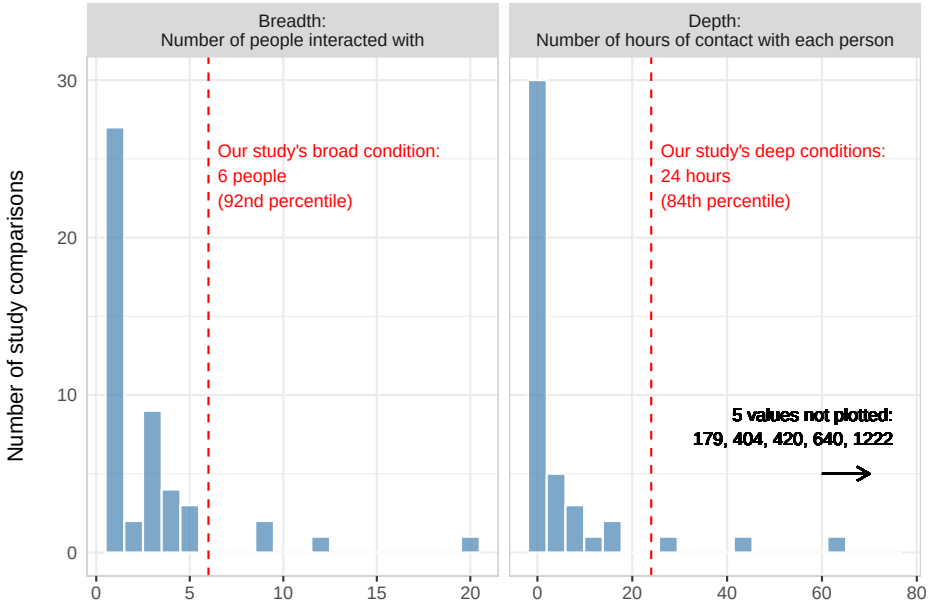
(a) Photograph of a set of bags roped up to be transported, in front of the entrance to the factory.

(b) Photograph of two workers collaborating on the production of a big bag inside a factory booth.

(c) Photograph of a completed small bag standing up.

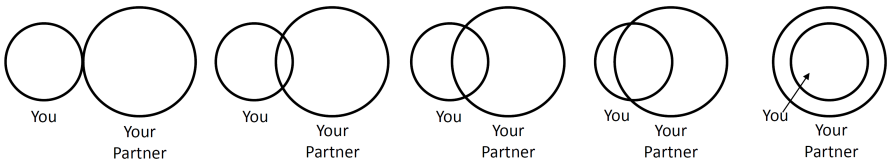
(d) Photograph of a completed small bag flattened.

Figure S2: Strength of our deep and broad treatments relative to the literature.



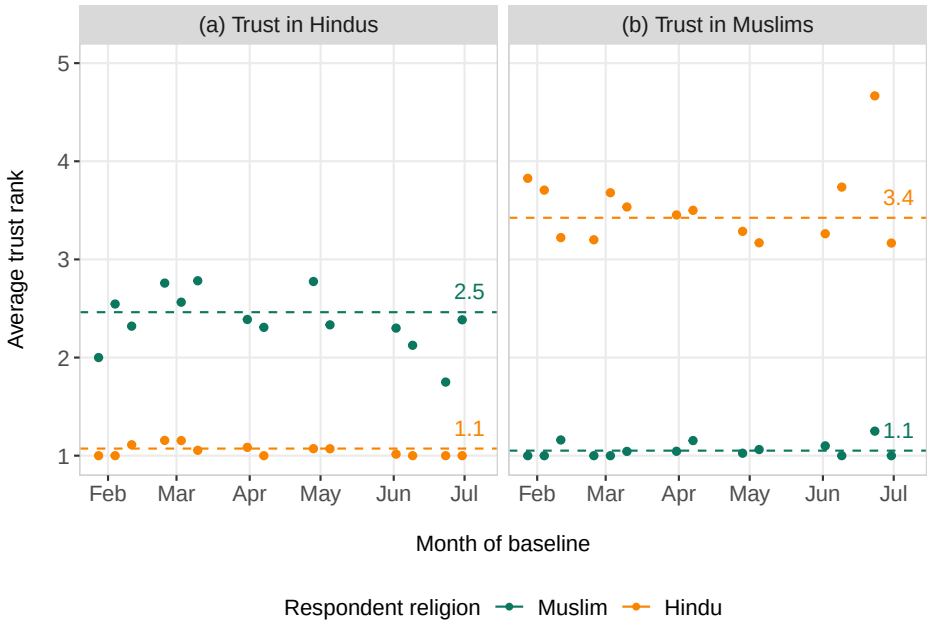

*Notes:* This figure presents where our treatments stand relative to pre-registered experiments on intergroup contact listed in the American Economic Association, the Evidence in Governance and Politics, and the Open Science Framework up to the end of December 2023. For treatment breadth, we compare our six-person interaction to the number of outgroup members interacted with in other studies, and for depth, we use the total hours of interaction per subject.

Figure S3: Inclusion of partner in self measure, from daily mood survey.



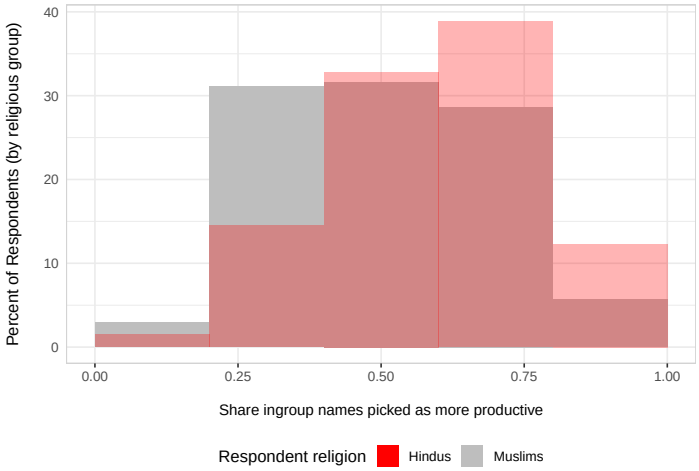

*Notes:* This figure shows the response options for the following question in the daily mood survey: "Which picture best describes your relationship with your partner today?"

Figure S4: Baseline trust rankings, by respondent religion.

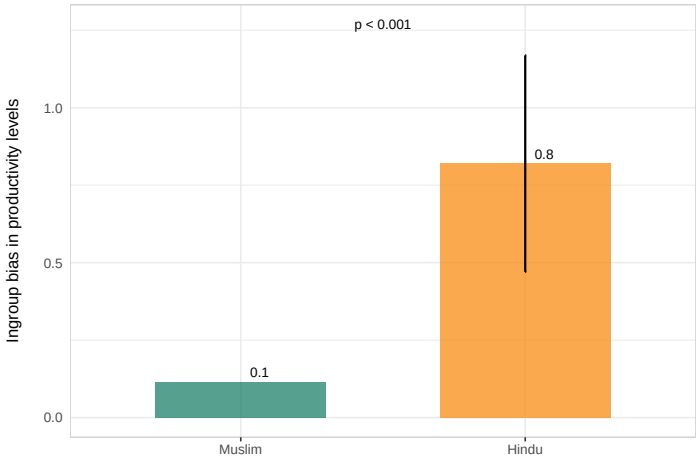



*Notes:* The dots in this figure show the average trust rankings, over two-week time intervals, for both Hindu and Muslim respondents, across the duration of the study. The data are from the Baseline 1 survey question: "Please rank the following groups in terms of how much you trust people from these groups. Rank 1 means you trust them the most and Rank 5 means you trust them the least." The ranking options (randomly ordered for each respondent) were: Muslims, Sikhs, People from north Indian states, People from Nepal, and Hindus. Dashed lines indicate the overall mean trust rankings for each group, with numeric labels showing these averages. The y-axis is scaled from 1 (most trust) to 5 (least trust).

Figure S5: Beliefs about the productivity of ingroup and outgroup members.



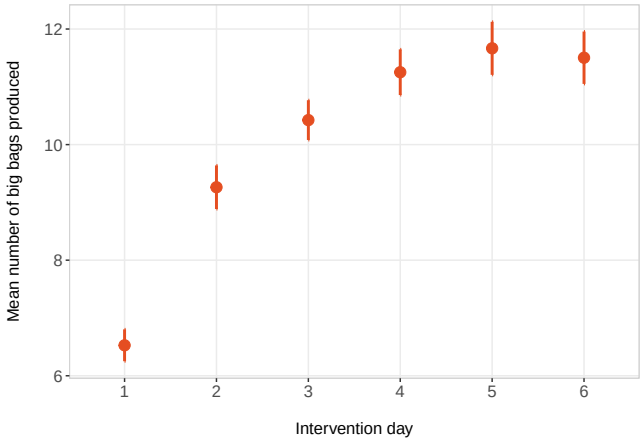

(a) Extensive margin.




(b) Combining extensive and intensive margins.

*Notes:* This figure presents results from the belief survey that was conducted with individuals that participated in the paper-bag making task. Each individual was first given five pairs of names (one same-religion pair and four mixed-religion pairs) and asked to predict which individual in the pair produced more paper bags per hour. Panel (a) shows the share of ingroup individuals reported as more productive by Hindus and Muslims. Respondents were then told the total number of bags produced by the pair per hour, and asked how many more bags their chosen individual in the previous question produced. Panel (b) presents the sum of intensive margin predictions where the ingroup member is predicted to be more productive, minus the sum of predictions for cases where the outgroup member is predicted to be more productive, divided by four. This provides the average additional production attributed to the ingroup member. 95% confidence intervals are based on robust standard errors.

Figure S6: Number of big bags produced by pair, by intervention day.

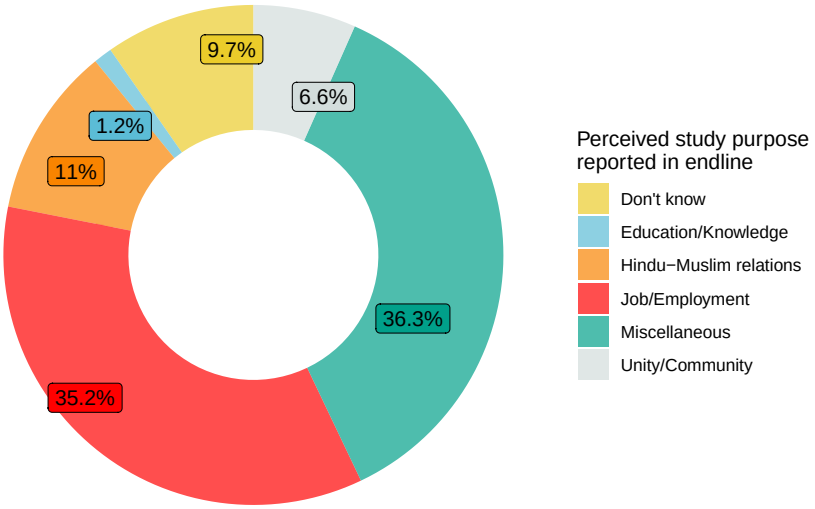



*Notes:* This figure shows the average number of big bags made across intervention days. Estimates are based on a regression of the number of big bags produced on indicator variables for intervention day and demeaned wave fixed effects. The unit of analysis is the worker-pair-by-day. 95% confidence intervals are based on standard errors clustered by unique worker-pair.

Figure S7: Exposure to other religious groups at work and in home ward, by respondent religion.

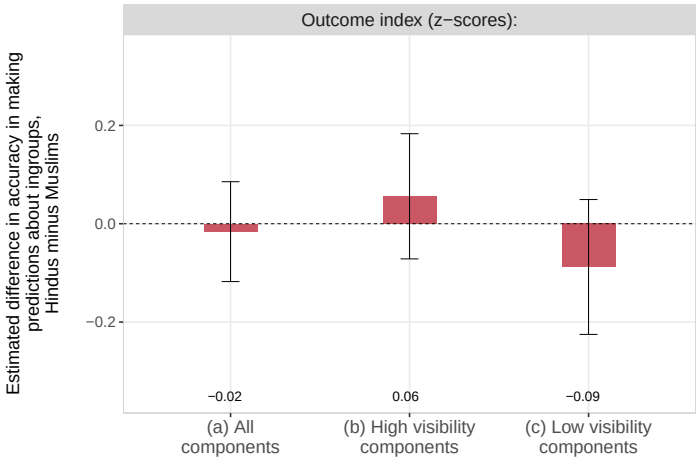
*Notes:* This figure summarizes how many people from different groups respondents estimate there are in the places where they live and work. The survey questions are: (a) "Think about your current place of work or study—or if you are not currently working or studying, your last place of work or study. Out of every 100 people there, how many belong(ed) to each of the following groups?" (b) "In your ward, out of every 100 people, roughly how many people belong to each of the following groups?" For obfuscation, subjects were asked to make these assessments for five groups: non-Bengalis, Muslims, people who own a scooter, Hindus, and people under the age of 18 (randomly ordered by respondent).

Figure S8: The purpose of the study, according to endline respondents.



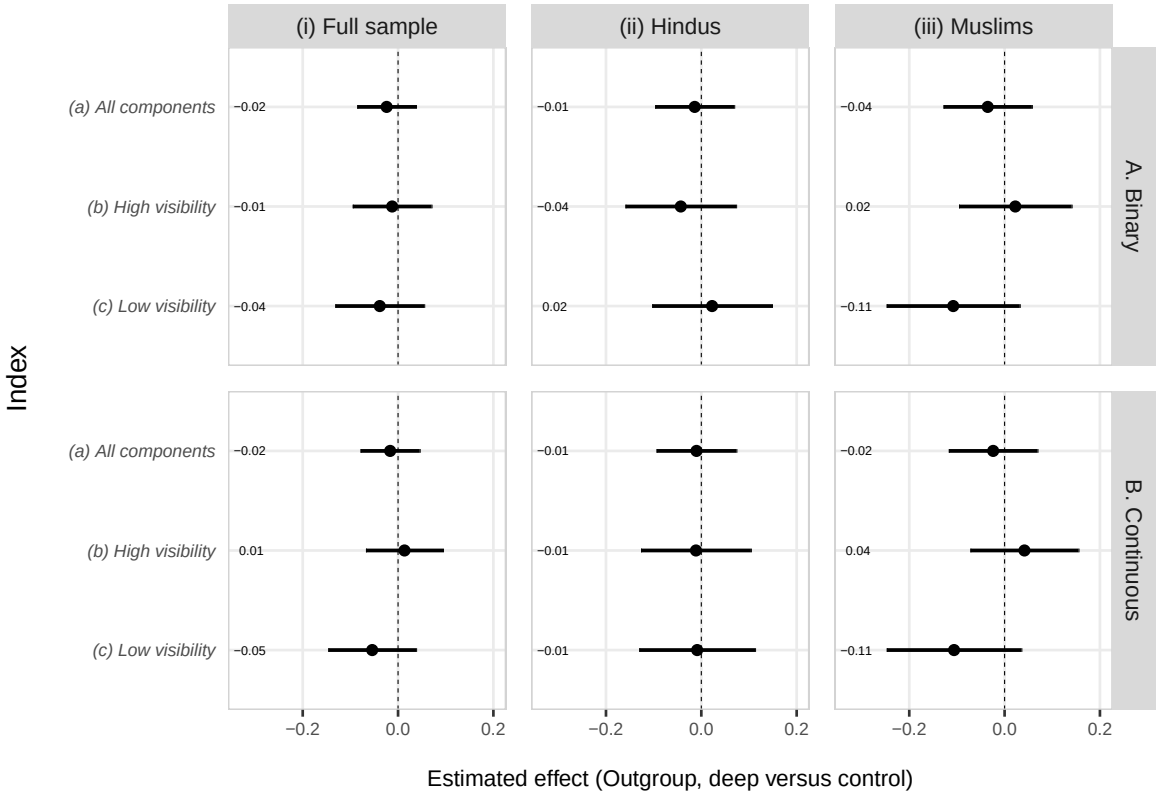

*Notes:* This figure summarizes endline subjects' reports of the purpose of the study. Subjects were asked at endline, "What do you think was the purpose of this research study? What was it trying to understand?" Subjects' qualitative responses were recorded by enumerators and categorized using ChatGPT5.

Figure S9: Control group Hindus make no more nor less accurate predictions about the ingroup than Muslims.



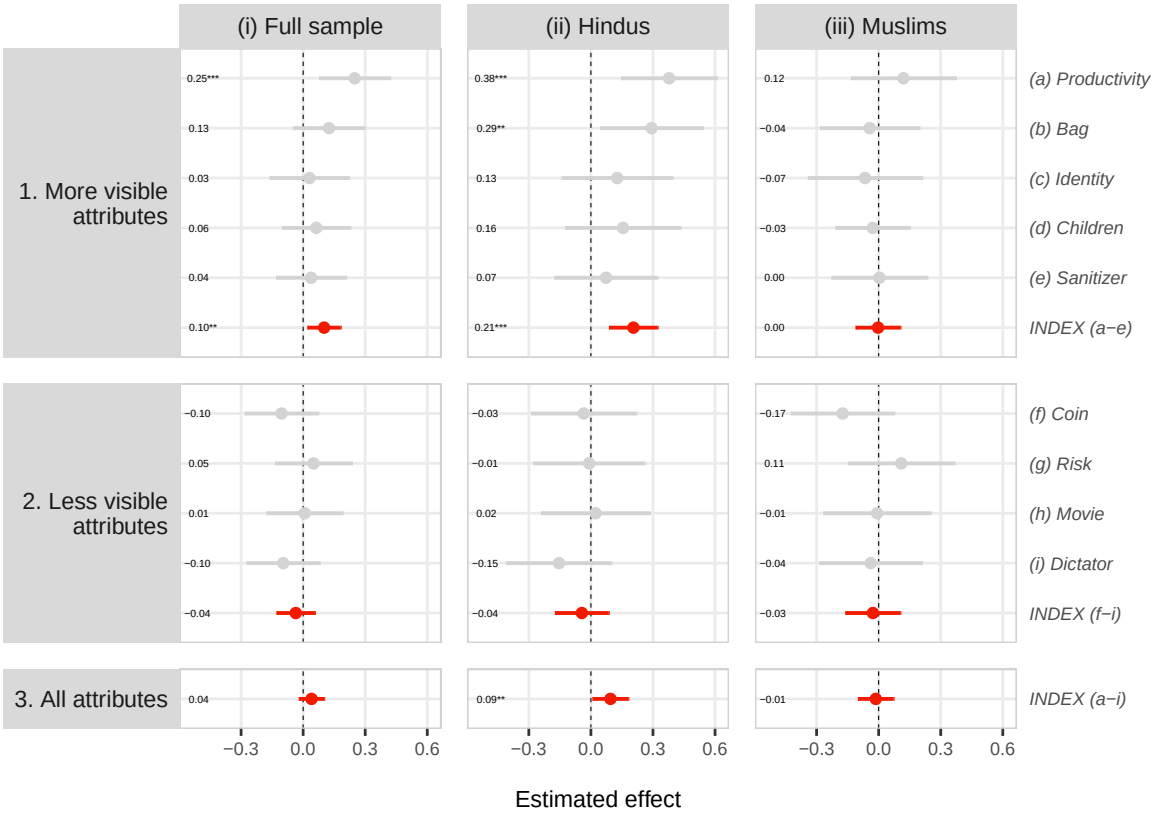

*Notes:* This figure parallels Figure 4, but now showing the difference in average accuracy, in standard deviations, between Hindus and Muslims in making predictions about ingroups, using only control group participants. In column (a), all eight components on which subjects made guesses about ingroups in the holdout sample (or about factory workers for productivity guesses), are standardized and averaged in a single index. Column (b) uses only the four high-visibility components: productivity, identity, children, and sanitizer (note, bag was not asked for ingroups). Column (c) uses only the remaining four low-visibility components: coin, risk, movie, and dictator. *p<0.1; **p<0.05; ***p<0.01.

Figure S10: *Outgroup, deep* does not increase prediction accuracy about the preferences and behaviors of outgroup strangers relative to the control group.



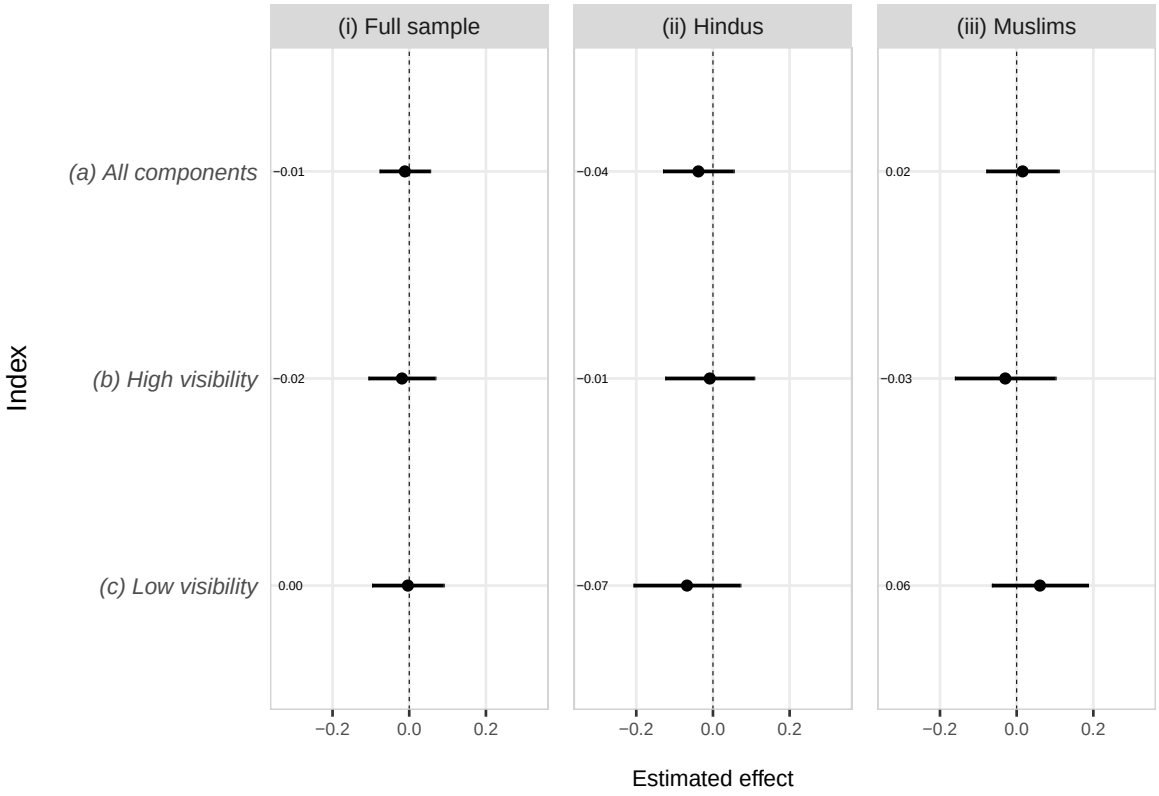

*Notes:* This figure parallels Figure 5, but shows the effect of *Outgroup, deep* on the accuracy of guesses about outgroups relative to the pure control group. Each dot and bar summarizes the result of one regression. The sample includes subjects assigned to *Outgroup, deep* or control. The unit of analysis is the subject. All outcomes are $z$-scores, centered and standardized according to the variable mean and standard deviation in the pure control condition. The (a) outcome ("All components") is the mean of the nine components measuring prediction accuracy, as described in the main text. The (b) outcome ("High visibility") is the mean of the five components: production, identity, bag, children, and sanitizer. The (c) outcome ("Low visibility") is the mean of the remaining four components: coin, risk, movie, and dictator. In panel A, all components (prior to standardization) are binary indicators for prediction accuracy. In panel B, the raw production, children, dictator, and coin components are treated as continuous; we take the absolute difference between the guess and "truth" benchmark, we add the integer one, we take the log of the resulting value (so that closer guesses receive more weight), and we then center and standardize the variable according to mean and standard deviation of the variable in the control group. Regressions include randomization strata fixed effects. The 95% confidence intervals and $p$-values for the differences are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

Figure S11: *Outgroup, broad* increases prediction accuracy about the more visible preferences and behaviors of outgroup strangers: Estimates for individual components.




*Notes:* This figure shows estimated treatment effects (in standard deviations) on the accuracy of predictions made about outgroups in the holdout sample. The unit of analysis is the subject. The sample includes subjects in the *Outgroup, deep* and *Outgroup, broad* conditions. All outcomes are $z$-scores, centered and standardized according to the variable mean and standard deviation in *Outgroup, deep*. The outcome for panels in row (3) is the mean of the nine components measuring prediction accuracy, as described in the main text. The outcome for panels in row (1) is the mean of the five components: productivity, bag, identity, children, and sanitizer. The outcome for panels in row (2) is the mean of the remaining four components: coin, risk, movie, and dictator. Regressions include randomization strata fixed effects. The 95% confidence intervals and $p$-values for the differences are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

Figure S12: *Outgroup, broad* does not increase prediction accuracy about the preferences and behaviors of ingroup strangers (placebo test).




*Notes:* This figure parallels Figure 5, but now shows estimated treatment effects on the accuracy of predictions made about *ingroups* in the holdout sample. The unit of analysis is the subject. All outcomes are averages of component $z$-scores, centered and standardized according to the variable mean and standard deviation in the *Outgroup, deep* condition. The outcome for panels in row (a) is the mean of the eight components measuring prediction accuracy, as described in the main text. (Note, the *Bag* component cannot be included in the indexes analyzed in this figure as this question was not asked about ingroups at endline.) The outcome for panels in row (b) is the mean of the four components: production, identity, children, and sanitizer. The outcome for panels in row (c) is the mean of the remaining four components: coin, risk, movie, and dictator. Regressions include randomization strata fixed effects. The 95% confidence intervals and $p$-values for the differences are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

# C Supplementary Text

## C.1 Relation to Pre-Registration

We pre-registered the experiment in the AEA RCT Registry under ID #13126, available at https://www.socialscienceregistry.org/trials/13126. We exhaustively report treatment effects on pre-registered primary outcomes. For our pre-registered secondary outcomes, we do not report an exhaustive analysis of these outcomes. Otherwise, we deviate from the pre-registration in the following ways:

- We do not report estimates of the pooled effect of contact, or of the pooled effect of

work relative to pure control (though the results of these comparisons can be to some extent inferred from the regressions we do report).

- We specified that our WhatsApp audit outcome would be "whether or not participants provided a polite response to the message." In practice, impolite responses were rare, and so we define the outcome instead as whether a response was received at all.

Finally, we note that we did not pre-specify the two dimensions of heterogeneity that we explore in the paper: Hindu versus Muslim, and high- versus low-visibility attributes. These results help us interpret our findings ex post through the lens of our formal model.

### C.2 Expression of $\widehat{x_{n,D}}^{gen}$ used in Equation 5

As defined previously, the inaccuracy $\Delta$ when guessing the attribute for person $i$ is defined as the expected squared distance between $\widehat{x_{n,D}}$ and the truth, that is $\Delta(\widehat{x_{n,D}}) = E(\widehat{x_{n,D}} - x^* - \eta_i)^2$. The accuracy $\mathscr{A}$ is just the negative of inaccuracy: $\mathscr{A}(\widehat{x_{n,D}}) = -\Delta(\widehat{x_{n,D}})$. We will use $\Delta^{gen}$ to denote the inaccuracy when guessing about a stranger $m \notin \{1, 2, \cdots, n\}$ and $\Delta^I$ for the inaccuracy when guessing about a previously-encountered partner $I \in \{1, 2, \cdots, n\}$.

$$
\begin{aligned}
\Delta^{gen} &= E\left(\widehat{x}(w^{st}, (w_{id})_{i,d}) - x^* - \eta_m\right)^2 \\
&= E\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^* - \eta_m\right)^2 \\
&= E\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^*\right)^2 \\
&\quad - 2E\left(\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^*\right)\eta_m\right) + E(\eta_m^2) \\
&= E\left(w^{st}(x_0^{st} - x^*) + \sum_{i\le n}\sum_{d\le D} w_{id}(x_{id} - x^*)\right)^2 + 0 + \sigma_{het}^2 \\
&= E\left(w^{st}(e^{st}) + \sum_{i\le n}\sum_{d\le D} w_{id}(\eta_i + \varepsilon_{id})\right)^2 + \sigma_{het}^2 \\
&= (w^{st})^2 E(e^{st})^2 + \sum_{i\le n}\sum_{d\le D}(w_{id})^2 E(\eta_i^2 + \varepsilon_{id}^2) + \sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})E(\eta_i^2) + \sigma_{het}^2 \\
&= (w^{st})^2\sigma_{st}^2 + \sum_{i\le n}\sum_{d\le D}(w_{id})^2(\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})\sigma_{het}^2 + \sigma_{het}^2 \qquad (7)
\end{aligned}
$$

$$= (1 - \sum_{i \le n} \sum_{d \le D} w_{id})^2 \sigma_{st}^2 + \sum_{i \le n} \sum_{d \le D} (w_{id})^2 (\sigma_{het}^2 + \sigma_{inf}^2)$$
$$+ \sum_{i \le n} \sum_{d > d'} (2 w_{id} w_{id'}) \sigma_{het}^2 + \sigma_{het}^2 \tag{8}$$

Here, the sum $\sum_{d>d'}$ includes each distinct pair of occasions once.

To derive the expression in (7), we use the statistical property $E(X - \bar{X})(Y - \bar{Y}) = EXY - E\bar{X}\bar{Y}$ and the assumptions about conditional means in Section 4, repeatedly for each pair of random variables. The first derivative of (8) with respect to any $w_{i^* d^*}$ is:

$$-2(1 - \sum_{i \le n} \sum_{d \le D} w_{id}) \sigma_{st}^2 + 2 w_{i^* d^*} (\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{d \ne d^*} (2 w_{i^* d}) \sigma_{het}^2 = 0$$

Let $J_k = \mathbf{1}_k \mathbf{1}_k^\top$. After eliminating $w^{st}$, the Hessian is

$$H = 2\sigma_{inf}^2 I_{nD} + 2\sigma_{het}^2 \operatorname{diag}(J_D, \ldots, J_D) + 2\sigma_{st}^2 J_{nD}.$$

This matrix is positive definite because $\sigma_{inf}^2 > 0$. The objective therefore has a unique global minimum. Its invariance to permutations of individuals and to permutations of occasions within each individual implies equal observation weights. Writing $w_{id} = w$, we obtain

$$-2(1 - \sum_{i \le n} \sum_{d \le D} w) \sigma_{st}^2 + 2 w_{i^* d^*} (\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{d \ne d^*} (2w) \sigma_{het}^2 = 0$$
$$\iff -(1 - nDw)\sigma_{st}^2 + w(\sigma_{het}^2 + \sigma_{inf}^2) + (D-1)(w)\sigma_{het}^2 = 0$$
$$\iff w(nD\sigma_{st}^2 + (\sigma_{het}^2 + \sigma_{inf}^2) + (D-1)\sigma_{het}^2) = \sigma_{st}^2$$

$$\iff w = \frac{\sigma_{st}^2}{(nD\sigma_{st}^2 + (\sigma_{het}^2 + \sigma_{inf}^2) + (D-1)\sigma_{het}^2)}$$
$$= \frac{\sigma_{st}^2}{(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)}$$
$$= \frac{\frac{1}{\sigma_{inf}^2 + D\sigma_{het}^2}}{(\frac{nD}{\sigma_{inf}^2 + D\sigma_{het}^2} + \frac{1}{\sigma_{st}^2})}$$
$$\iff w^{st} = 1 - nDw = \frac{\frac{1}{\sigma_{st}^2}}{(\frac{nD}{\sigma_{inf}^2 + D\sigma_{het}^2} + \frac{1}{\sigma_{st}^2})}$$

### C.3 Expression of $\widehat{x_{n,D}}^I$ used in Equation 6

Let $w_1$ denote the common weight placed on each observation from an encountered individual $i \neq I$, and let $w_2$ denote the common weight placed on each observation from individual $I$. By symmetry, observations within each of these two classes receive the same weight.

The weights satisfy

$$w^{st} + (n-1)Dw_1 + Dw_2 = 1.$$

From the DM's perspective, the inaccuracy of the estimate of the encountered individual $I$ can be written as

$$\begin{aligned}\Delta^I = &(w^{st})^2\sigma_{st}^2 + (Dw_2 - 1)^2\sigma_{het}^2 \\ &+ (n-1)D^2 w_1^2 \sigma_{het}^2 + \left[(n-1)Dw_1^2 + Dw_2^2\right]\sigma_{inf}^2.\end{aligned}$$

The first term is the contribution of uncertainty in the prior. The second captures uncertainty about the individual-specific component $\eta_I$ of the encountered partner. The third captures heterogeneity across the other $n-1$ encountered individuals, and the final term is occasion-specific observation noise.

Substituting

$$w^{st} = 1 - (n-1)Dw_1 - Dw_2$$

and minimizing with respect to $w_1$ and $w_2$ gives

$$w_1 = \frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2 + D\sigma_{het}^2)(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)},$$

$$w_2 = \frac{\sigma_{st}^2\sigma_{inf}^2 + nD\sigma_{st}^2\sigma_{het}^2 + \sigma_{inf}^2\sigma_{het}^2 + D\sigma_{het}^4}{(\sigma_{inf}^2 + D\sigma_{het}^2)(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)},$$

and

$$w^{st} = \frac{\sigma_{inf}^2}{nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2}.$$

### C.4 Omitted proof for Result 1

*Proof.* Part (i): In $\Delta^{gen}$ derived in Equation 7, we change $E(x_0^{st}-x^*)^2$ from $E(e^{st})^2=\sigma_{st}^2$ to $(x_0^{st}-x^*)^2$, to get:

$$
\begin{aligned}
\Delta^{gen} &= (w^{st})^2(x_0^{st}-x^*)^2+\sum_{i\le n}\sum_{d\le D}(w_{id})^2E(\eta_i^2+\varepsilon_{id}^2)+\sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})E(\eta_i^2)+\sigma_{het}^2\\
&= \left(\frac{\sigma_{inf}^2+D\sigma_{het}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2(x_0^{st}-x^*)^2\\
&\quad +nD\left(\frac{\sigma_{st}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2(\sigma_{het}^2+\sigma_{inf}^2)\\
&\quad +nD(D-1)\left(\frac{\sigma_{st}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2\sigma_{het}^2+\sigma_{het}^2\\
&= \frac{(\sigma_{inf}^2+D\sigma_{het}^2)^2b^2+nD\sigma_{st}^4\left((\sigma_{het}^2+\sigma_{inf}^2)+(D-1)\sigma_{het}^2\right)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}+\sigma_{het}^2\\
&= (\sigma_{inf}^2+D\sigma_{het}^2)\frac{(\sigma_{inf}^2+D\sigma_{het}^2)b^2+nD\sigma_{st}^4}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}+\sigma_{het}^2
\end{aligned}
$$

$$
\frac{\partial\Delta^{gen}}{\partial n}=-(\sigma_{inf}^2+D\sigma_{het}^2)\frac{D\sigma_{st}^2\left(2b^2\sigma_{inf}^2+2Db^2\sigma_{het}^2+nD\sigma_{st}^4-\sigma_{st}^2\sigma_{inf}^2-D\sigma_{st}^2\sigma_{het}^2\right)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^3}
$$

$\frac{\partial\Delta^{gen}}{\partial n}<0$ is guaranteed by $2b^2\ge\sigma_{st}^2$ and hence also by $3b^2\ge 2\sigma_{st}^2$.

$$
\frac{\partial^2\Delta^{gen}}{\partial n^2}=(\sigma_{inf}^2+D\sigma_{het}^2)\frac{2D^2\sigma_{st}^4\left(3b^2\sigma_{inf}^2+3Db^2\sigma_{het}^2+nD\sigma_{st}^4-2\sigma_{st}^2\sigma_{inf}^2-2D\sigma_{st}^2\sigma_{het}^2\right)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^4}
$$

$\frac{d^2\Delta^{gen}}{dn^2}>0$ is guaranteed by $3b^2\ge 2\sigma_{st}^2$.

Next, we calculate the comparative statics with respect to $D$. Recall that

$$
\Delta^{gen}=\sigma_{het}^2+(\sigma_{inf}^2+D\sigma_{het}^2)\frac{(\sigma_{inf}^2+D\sigma_{het}^2)b^2+nD\sigma_{st}^4}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}.
$$

Taking the derivative with respect to $D$ gives

$$\frac{\partial \Delta^{gen}}{\partial D} = \frac{n\sigma_{st}^2\sigma_{inf}^2\left[(\sigma_{st}^2 - 2b^2)(\sigma_{inf}^2 + D\sigma_{het}^2) - nD\sigma_{st}^4\right]}{(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)^3}.$$

Hence $b^2 \geq \frac{1}{2}\sigma_{st}^2$ is sufficient for $\frac{\partial \Delta^{gen}}{\partial D} < 0$. In particular, this is implied by our maintained assumption $b^2 \geq \frac{2}{3}\sigma_{st}^2$.

To establish diminishing returns, let

$$q = b^2 - \frac{2}{3}\sigma_{st}^2 \geq 0.$$

Then

$$\frac{\partial^2 \Delta^{gen}}{\partial D^2} = \frac{2n\sigma_{st}^2\sigma_{inf}^2}{(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)^4}\begin{bmatrix} \frac{1}{3}D\sigma_{het}^4\sigma_{st}^2 + \frac{4}{3}Dn\sigma_{het}^2\sigma_{st}^4 \\ \quad + Dn^2\sigma_{st}^6 + \frac{1}{3}\sigma_{het}^2\sigma_{st}^2\sigma_{inf}^2 \\ \quad + q\big(2D\sigma_{het}^4 + 2Dn\sigma_{het}^2\sigma_{st}^2 \\ \qquad + 2\sigma_{het}^2\sigma_{inf}^2 + 3n\sigma_{st}^2\sigma_{inf}^2\big) \end{bmatrix} > 0.$$

Thus, under $b^2 \geq \frac{2}{3}\sigma_{st}^2$, increasing $D$ reduces inaccuracy with diminishing returns. No additional condition on $nD^2\sigma_{st}^2$ is required.

ii) Next, we fix $nD = N$ and take derivatives w.r.t $D$.

$$\frac{\partial \Delta^{gen}}{\partial D}|_{N=nD} = \frac{N\sigma_{st}^2\sigma_{het}^2\left(2b^2\sigma_{inf}^2 + 2Db^2\sigma_{het}^2 + N\sigma_{st}^4 - \sigma_{st}^2\sigma_{inf}^2 - D\sigma_{st}^2\sigma_{het}^2\right)}{(N\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)^3}$$

$\frac{\partial \Delta^{gen}}{\partial D}|_{N=nD} > 0$ is guaranteed by $2b^2 \geq \sigma_{st}^2$ and hence also by $3b^2 \geq 2\sigma_{st}^2$.

For the analyst's evaluation of accuracy, let $b = |x_0^{st} - x^*|$ denote the realized error in the initial belief. Substituting the optimal weights from Equation 6 into the expression for $\Delta^I$ gives

$$\Delta^I = \frac{\sigma_{het}^2\sigma_{inf}^2}{\sigma_{inf}^2 + D\sigma_{het}^2} + \frac{\sigma_{st}^2\sigma_{inf}^4}{(\sigma_{inf}^2 + D\sigma_{het}^2)(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)} + \frac{(b^2 - \sigma_{st}^2)\sigma_{inf}^4}{(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)^2}.$$

Now fix total contact at $N = nD$, and define

$$A = \sigma_{inf}^2 + D\sigma_{het}^2, \qquad C = N\sigma_{st}^2 + A.$$

Then

$$\Delta^I = \frac{\sigma_{het}^2 \sigma_{inf}^2}{A} + \frac{\sigma_{st}^2 \sigma_{inf}^4}{AC} + \frac{(b^2 - \sigma_{st}^2)\sigma_{inf}^4}{C^2},$$

so

$$\left.\frac{d\Delta^I}{dD}\right|_{N=nD} = -\frac{\sigma_{het}^4 \sigma_{inf}^2}{A^2} - \frac{\sigma_{st}^2 \sigma_{het}^2 \sigma_{inf}^4 (A+C)}{A^2 C^2} - \frac{2\sigma_{het}^2 (b^2 - \sigma_{st}^2)\sigma_{inf}^4}{C^3}.$$

If $b^2 \geq \sigma_{st}^2$, every term on the right-hand side is non-positive and the first is strictly negative.

If $b^2 < \sigma_{st}^2$, the final term is positive, but it is bounded above by

$$\frac{2\sigma_{st}^2 \sigma_{het}^2 \sigma_{inf}^4}{C^3}.$$

Since $C \geq A$,

$$\frac{A+C}{A^2 C^2} \geq \frac{2}{C^3},$$

so the magnitude of the second negative term is at least as large as this upper bound. The first term remains strictly negative. Therefore,

$$\left.\frac{d\Delta^I}{dD}\right|_{N=nD} < 0.$$

Hence, holding total contact $N = nD$ fixed, greater depth strictly improves accuracy about the encountered individual.

**Part (iv).** Equation (5) implies

$$\widehat{x_{n,D}}^{gen} = w^{st} x_0^{st} + (1 - w^{st})\bar{x}.$$

Substituting $D = N/n$ gives the expression in part (iv). Its derivative with respect to $n$, holding $N$ fixed, is

$$\left.\frac{d(1-w^{st})}{dn}\right|_N = \frac{N^2\sigma_{st}^2\sigma_{het}^2}{n^2\left(N\sigma_{st}^2+\sigma_{inf}^2+\frac{N}{n}\sigma_{het}^2\right)^2} > 0.$$

□

### C.5 Bayesian updating under normality

In deriving Result 1, we modeled the DM as a frequentist, treating the true outgroup average as a fixed unknown constant that can be learned from the observed data using an optimal linear estimator. In this section, we show that modeling the DM as a Bayesian—who begins with an explicit prior probability distribution over the unknown group average and updates to a posterior distribution—would still lead to identical conclusions as in Equations 5 and Equations 6.

The Bayesian paradigm requires two additional distributional assumptions, about the normality of the prior belief, and the normality of all the random variables $\eta$ and $\varepsilon$, which leads to the joint normality of observed signals.

Specifically, suppose the DM holds the prior belief-distribution $x^* \sim N(x_0^{st}, \sigma_{st}^2)$ over the unknown constant $x^*$, which is the average trait of the outgroup. Note that as before, the DM knows $x_0^{st}$ and not $x^*$, but now $x^* = x_0^{st} + e^{st}$. We assume that $x^*$, all $\eta_i$, and all $\varepsilon_{id}$ are mutually independent, with $\eta_i \sim N(0, \sigma_{het}^2)$ and $\varepsilon_{id} \sim N(0, \sigma_{inf}^2)$.

Let $E$ denote the expectation with respect to the joint distribution of all $e^{st}, \varepsilon, \eta$ variables, for *all* individuals including $m$. Let $X^* = E[X \mid x^*] = x^*\mathbf{1}_{nD}$. Let $\mathbf{x}^*$ be a $D \times 1$ vector each of whose entries is $x^*$.

Let $p$ be a probability defined over the joint event space of all uncertainty pertaining to all random variables $\varepsilon, \eta$ and $x^*$. The normality and independence assumptions imply that conditional on $x^*$, the observed $nD \times 1$ information-vector

$$X = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})^T = (x_{11}, x_{12}, .., , x_{1D}, ..., x_{n1}, x_{n2}, .., , x_{nD})^T$$

is jointly normally distributed as $p(X|x^*) = N(X^*, \Sigma)$, where $\Sigma = E\left[(X-X^*)(X-X^*)^\top \mid x^*\right]$ is derived from the assumptions we made about the DM's model of the world in Section 4.1.

As a proof of concept, in Result 2, we focus only on the case of updating beliefs about a stranger from the outgroup. The other case, involving an encountered outgroup member, is mathematically similar, and we provide a proof outline after Result 2.

**Result 2.** *Given the assumptions above, a Bayesian DM's posterior beliefs $p(x^*|X)$ and $p(x^* + \eta_m|X)$ are both normally distributed with mean equal to*

$$\mu_{gen} = \frac{1}{\frac{1}{\sigma_{st}^2} + \frac{n}{\sigma_{inf}^2/D + \sigma_{het}^2}} \left( \frac{1}{\sigma_{st}^2} \times x_0^{st} + \sum_{i \leq n} \frac{1}{\sigma_{inf}^2/D + \sigma_{het}^2} \times \frac{\sum_{d \leq D} x_{id}}{D} \right).$$

*Proof.* We would be using the following property of jointly normal variables,

$$p(X|x^*) \propto \exp\left( -\frac{1}{2} (X - X^*)^{\mathrm{T}} \Sigma^{-1} (X - X^*) \right)$$

computing which requires inverting the matrix $\Sigma$.

As all observations across individuals are independent conditional on $x^*$, $\Sigma$ is a block-diagonal matrix, where every main-diagonal block is always the same $D \times D$ matrix $M$ and all off-diagonal blocks are zero matrices. Within $M$, all diagonal elements are $\sigma_{het}^2 + \sigma_{inf}^2$, and off-diagonal elements are $\sigma_{het}^2$.

$$\Sigma = \begin{pmatrix} M & 0 & 0 & 0... \\ 0 & M & 0 & 0... \\ 0 & 0 & M & 0... \\ 0 & 0 & 0 & M... \\ ... & ... & ... & ... \end{pmatrix}_{nD \times nD}$$

$$\text{where } M = \begin{pmatrix} \sigma_{het}^2 + \sigma_{inf}^2 & \sigma_{het}^2 & \sigma_{het}^2 & ... \\ \sigma_{het}^2 & \sigma_{het}^2 + \sigma_{inf}^2 & \sigma_{het}^2 & ... \\ \sigma_{het}^2 & \sigma_{het}^2 & \sigma_{het}^2 + \sigma_{inf}^2 & ... \\ ... & ... & ... & ... \end{pmatrix}_{D \times D}$$

The DM's posterior according to Bayes rule is

$$\begin{aligned} p(x^*|X) &\propto P(X|x^*)P(x^*) \\ &\propto \exp\left( -\frac{1}{2} (X - X^*)^{\mathrm{T}} \Sigma^{-1} (X - X^*) \right) \exp\left( -\frac{1}{2} \frac{(x^* - x_0^{st})^2}{\sigma_{st}^2} \right) \end{aligned}$$

which essentially depends on calculating $\Sigma^{-1}$. Given the block diagonal structure, this requires inverting $M$. First, we describe an algorithm to invert the matrix $M$ (for arbitrary parameter values). This uses the Gauss-Jordan method, performing the following row operations on $M$:

$$\left( \sum_{i=1}^{D} R_i \right) \to R_D \implies \text{all entries in } R_D \text{ become } D\sigma_{het}^2 + \sigma_{inf}^2$$

$$\frac{1}{D\sigma_{het}^2+\sigma_{inf}^2}R_D \to R_D \implies \text{all entries in } R_D \text{ become } 1$$

$$\forall j \in \{1,2,3,...,D-1\},\ \left(R_j - \sigma_{het}^2 R_D\right) \to R_j \implies \text{all diagonal and off-diagonal elements in } R_j \text{ become } \sigma_{inf}^2, 0 \text{ respectively}$$

$$\forall j \in \{1,2,3,...,D-1\},\ \frac{1}{\sigma_{inf}^2}R_j \to R_j \implies \text{all diagonal and off-diagonal elements in } R_j \text{ become } 1, 0 \text{ respectively}$$

$$\left(R_D - \sum_{i=1}^{D-1} R_i\right) \to R_D \implies \text{all the diagonal and all off-diagonal elements in } M \text{ become } 1, 0 \text{ respectively}$$

These five row operations convert $M$ to an identity matrix $I$ of the same dimension $D \times D$. The same five row operations, when applied to the rows $\rho_1, ..., \rho_D$ of the identity matrix $I$, will transform it into $M^{-1}$, through the following stages of transformation:

1) $\rho_D$ becomes a row of all 1s.
2) $\rho_D$ becomes a row of all $\frac{1}{D\sigma_{het}^2+\sigma_{inf}^2}$ terms.
3) In $\rho_j : j < D$, all diagonal elements become $1 - \frac{\sigma_{het}^2}{D\sigma_{het}^2+\sigma_{inf}^2}$, off-diagonal elements become $-\frac{\sigma_{het}^2}{D\sigma_{het}^2+\sigma_{inf}^2}$.
4) In $\rho_j : j < D$, all diagonal elements become $\frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}$, off-diagonal elements become $-\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}$.
5) In $\rho_D$, the diagonal element becomes $\frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}$, off-diagonal elements become $-\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}$.

It follows that $\Sigma^{-1}$ is a block-diagonal matrix, where every main-diagonal block is always the same $D \times D$ matrix $M^{-1}$ and all off-diagonal blocks are zero matrices.

$$\Sigma^{-1} = \begin{pmatrix} M^{-1} & 0 & 0 & 0... \\ 0 & M^{-1} & 0 & 0... \\ 0 & 0 & M^{-1} & 0... \\ 0 & 0 & 0 & M^{-1}... \\ ... & ... & ... & ... \end{pmatrix}_{nD \times nD}$$

$$\text{where } M^{-1} = \begin{pmatrix} \frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & \cdots \\ -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & \frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & \cdots \\ -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & -\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & \frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}_{D\times D}$$

Now, due to the block diagonal structure of $\Sigma^{-1}$,

$$(X-X^*)^T\Sigma^{-1}(X-X^*) = \sum_{i=1}^{n}(\mathbf{x_i}-\mathbf{x}^*)^T M^{-1}(\mathbf{x_i}-\mathbf{x}^*)$$

Next, writing $\bar{x}_i = \sum_{d=1}^{D} x_{id}/D$

$$\begin{aligned}
&(\mathbf{x_i}-\mathbf{x}^*)^T M^{-1}(\mathbf{x_i}-\mathbf{x}^*) \\
&= (x^*)^2\left(D\times\frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)} + D(D-1)\times\left(-\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}\right)\right) \\
&\quad - 2x^*\sum_{d=1}^{D}x_{id}\left((D-1)\times\left(-\frac{\sigma_{het}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}\right)+\frac{(D-1)\sigma_{het}^2+\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}\right)+K \\
&= (x^*)^2\left(\frac{D\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}\right) - 2x^*\sum_{d=1}^{D}x_{id}\times\left(\frac{\sigma_{inf}^2}{\sigma_{inf}^2(D\sigma_{het}^2+\sigma_{inf}^2)}\right)+K \\
&= (x^*)^2\left(\frac{D}{(D\sigma_{het}^2+\sigma_{inf}^2)}\right) - 2x^*\sum_{d=1}^{D}x_{id}\times\left(\frac{1}{(D\sigma_{het}^2+\sigma_{inf}^2)}\right)+K \\
&= (x^*)^2\left(\frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)}\right) - 2x^*\bar{x}_i\times\left(\frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)}\right)+K
\end{aligned}$$

where $K$ is a constant that is independent of $x^*$. Therefore,

$$\begin{aligned}
p(x^*|X) &\propto \exp\left(-\frac{1}{2}(X-X^*)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(X-X^*)\right)\exp\left(-\frac{1}{2}\frac{(x^*-x_0^{st})^2}{\sigma_{st}^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x_i}-\mathbf{x}^*)^T M^{-1}(\mathbf{x_i}-\mathbf{x}^*)\right)\exp\left(-\frac{1}{2}\frac{(x^*-x_0^{st})^2}{\sigma_{st}^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left((x^*)^2\left(\frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)}\right)-2x^*\bar{x}_i\times\left(\frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)}\right)\right)\right)\exp\left(-\frac{1}{2}\frac{(x^*-x_0^{st})^2}{\sigma_{st}^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left((x^*)^2\left(\frac{n}{(\sigma_{het}^2+\sigma_{inf}^2/D)}+\frac{1}{\sigma_{st}^2}\right)-2x^*\left(\sum_{i=1}^{n}\bar{x}_i\times\frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)}+\frac{x_0^{st}}{\sigma_{st}^2}\right)\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&\propto \exp\left(-\frac{1}{2\sigma_1^2}\left(x^* - \frac{\left(\sum_{i=1}^n \bar{x}_i \times \frac{1}{(\sigma_{het}^2+\sigma_{inf}^2/D)} + \frac{x_0^{st}}{\sigma_{st}^2}\right)}{\left(\frac{n}{(\sigma_{het}^2+\sigma_{inf}^2/D)} + \frac{1}{\sigma_{st}^2}\right)}\right)^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma_1^2}(x^* - \mu_{gen})^2\right)
\end{aligned}
$$

where we have used the abbreviation $\sigma_1^2 = \frac{1}{\left(\frac{n}{(\sigma_{het}^2+\sigma_{inf}^2/D)} + \frac{1}{\sigma_{st}^2}\right)}$. Therefore $p(x^*|X)$ is normally distributed, centered at $\mu_{gen}$. Finally, using the fact that $\eta_m$ is independent of $X$ and $x^*$,

$$
\begin{aligned}
&p(x^* + \eta_m = y|X) \\
&= \int_{x^*} p(x^* = x^*, \eta_m = y - x^*|X)dx^* \\
&= \int_{x^*} p(x^* = x^*|X)p(\eta_m = y - x^*)dx^* \\
&\propto \int_{x^*} \exp\left(-\frac{1}{2\sigma_1^2}(x^* - \mu_{gen})^2\right)\exp\left(-\frac{1}{2\sigma_{het}^2}(y - x^*)^2\right)dx^* \\
&\propto \int_{x^*} \exp\Bigg(-\frac{1}{2\sigma_1^2\sigma_{het}^2}\big((x^*)^2\sigma_{het}^2 + \mu_{gen}^2\sigma_{het}^2 - 2x^*\mu_{gen}\sigma_{het}^2 \\
&\qquad\qquad + y^2\sigma_1^2 + (x^*)^2\sigma_1^2 - 2x^*y\sigma_1^2\big)\Bigg)dx^* \\
&\propto \int_{x^*} \exp\left(-\frac{\sigma_1^2+\sigma_{het}^2}{2\sigma_1^2\sigma_{het}^2}\left((x^*)^2 - 2x^*\frac{y\sigma_1^2 + \mu_{gen}\sigma_{het}^2}{\sigma_1^2+\sigma_{het}^2} + \frac{\mu_{gen}^2\sigma_{het}^2 + y^2\sigma_1^2}{\sigma_1^2+\sigma_{het}^2}\right)\right)dx^* \\
&\propto \int_{x^*} \exp\left(-\frac{\sigma_1^2+\sigma_{het}^2}{2\sigma_1^2\sigma_{het}^2}\left((x^* - \frac{y\sigma_1^2 + \mu_{gen}\sigma_{het}^2}{\sigma_1^2+\sigma_{het}^2})^2 + \sigma_1^2\sigma_{het}^2\frac{(y-\mu_{gen})^2}{(\sigma_1^2+\sigma_{het}^2)^2}\right)\right)dx^* \\
&\propto \exp\left(-\frac{(y-\mu_{gen})^2}{2(\sigma_1^2+\sigma_{het}^2)}\right)\underbrace{\int_{x^*} \exp\left(-\frac{\sigma_1^2+\sigma_{het}^2}{2\sigma_1^2\sigma_{het}^2}\left((x^* - \frac{y\sigma_1^2 + \mu_{gen}\sigma_{het}^2}{\sigma_1^2+\sigma_{het}^2})^2\right)\right)dx^*}_{\sqrt{\frac{2\pi\sigma_1^2\sigma_{het}^2}{\sigma_1^2+\sigma_{het}^2}}}
\end{aligned}
$$

Thus, the posterior about $x^* + \eta_m$ has a normal distribution, centered at $\mu_{gen}$. This concludes the proof. The last few steps constitute a standard convolution integral. The final integral is a Gaussian integral whose value is independent of $y$ and is therefore absorbed into the proportionality constant.

□

How would the proof change when calculating $p(x^* + \eta_I)$, the posterior about an

encountered member? Here is a brief outline: In this case, one instead starts with the $(nD+1)\times 1$ vector

$$X = \left(x^* + \eta_I, x_I, x_1, \ldots, x_{I-1}, x_{I+1}, \ldots, x_n\right)^T.$$

Including $x^* + \eta_I$ in the vector and rearranging the vector to place $\mathbf{x_I}$ right next to it, helps account for the correlation between what the DM needs to form their posterior about and what he has already observed ($\mathbf{x_I}$). It also helps in calculating the inverse more easily. Redefine $X^* = E(X|x^*)$.

$\Sigma = E\left[(X-X^*)(X-X^*)^\top \mid x^*\right]$ would now be an $(nD+1)\times(nD+1)$ matrix,

$$\Sigma = \begin{pmatrix} M' & 0 & 0 & 0 & 0\ldots \\ 0 & M & 0 & 0 & 0\ldots \\ 0 & 0 & M & 0 & 0\ldots \\ 0 & 0 & 0 & M & 0\ldots \\ 0 & 0 & 0 & 0 & M\ldots \\ 0 & \ldots & \ldots & \ldots & \ldots \end{pmatrix}_{(nD+1)\times(nD+1)} \qquad \text{where } M' = \begin{pmatrix} \sigma^2_{het} & \sigma^2_{het} \\ \sigma^2_{het} & M \end{pmatrix}_{(D+1)\times(D+1)}$$

and $M$ stays the same as in the last proof. $\Sigma^{-1}$ is calculated in the same way as before, using the Gauss–Jordan method, but this time also calculating $(M')^{-1}$. Let $X_{-1}$ be the last $nD$ components of the $X$-vector, i.e., the DM's observations.

$$\begin{aligned} p(x^* + \eta_I = y|X_{-1}) &\propto \int_{x^*} p(x^* + \eta_I = y, X_{-1}|x^*)p(x^*)dx^* \\ &\propto \int_{x^*} \exp\left(-\frac{1}{2}(X-X^*)^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\,(X-X^*)\right)p(x^*)dx^* \end{aligned}$$

This expression is in a convolution integral form as before, though one requires quite a bit more algebra to get to the final expression. We skip this for brevity.

### C.6 Research ethics

We took a number of steps to ensure that the study conformed to commonly shared values governing research ethics.

- All participants in the study were adults aged 18 or over who provided fully informed consent in order to participate in the study. No deception was employed at any stage of the study.

- Individuals' work and time were well-compensated, beyond the local minimum wage, although not to a level that could reasonably be considered coercive.
- One issue with factory settings is the potential for industrial hazards. The production task was carefully selected to ensure that it involved no foreseeable dangers to workers. The only materials needed for the task were newspaper, a ruler, flour, water, and salt. Working conditions were kept as comfortable as possible. Workers were paid fixed salaries rather than piece rates, and no significant pressure was placed on them in the production process (in part to allow space for discussion and friend-making).
- The factory was not set up to make a profit and we did not sell the bags being produced—so as not to harm the businesses and livelihoods of local bag-makers. Instead, we donated the bags to a local paper vendor for pulping. However, to maintain motivation and naturalism, subjects were not alerted to the fact that the bags were not being sold for use.
- We anticipated substantial social learning due to the study. The results can inform optimal ways to structure intergroup interactions in a range of settings, from workplaces to schools and sports leagues.
- The sample size was only as large as we deemed necessary ex ante to estimate reasonably sized treatment effects.
- The study protocols were separately reviewed and approved by four Institutional Review Boards, at the University of California San Diego, the University of California Davis, Duke University, and the University of British Columbia.

### C.7 WhatsApp measure

For the WhatsApp measure, which was fielded approximately one month after the final endline survey, each subject was sent one of the text messages below (assigned at random) through WhatsApp. The text message came from a real enumerator: someone working with our survey firm, but who had not previously been involved with our study, and so was unknown to participants. Messages were sent in Bengali. All subjects received a message from an enumerator who belonged to the religious outgroup.

- Hi, I'm [HINDU NAME/MUSLIM NAME] from the bag factory. How is everything going?
- Hi, this is [HINDU NAME/MUSLIM NAME]. I'm from the bag factory team. Just checking in. Is everyone doing well at home?

- Hello, my name is [HINDU NAME/MUSLIM NAME], and I'm working with the bag factory project. All good with you and your family these days?
- Hi, this is [HINDU NAME/MUSLIM NAME] from the bag factory. Just following up. Are you and your family doing well?
- Hi, I'm [HINDU NAME/MUSLIM NAME] from the bag factory. How's everything going with you and your family?
- Hi, this is [HINDU NAME/MUSLIM NAME] from the bag factory team. How are things at home with you and family?
- Hello, my name is [HINDU NAME/MUSLIM NAME] from the bag factory project. All well with you and your family?
- Hi, I'm [HINDU NAME/MUSLIM NAME] from the bag factory. How's everyone doing at home with your family and you?
- Hi, this is [HINDU NAME/MUSLIM NAME] from the bag factory team. Just wanted to check in and see how you're doing. All well at home?
- Hello, my name is [HINDU NAME/MUSLIM NAME], and I'm with the bag factory project. Is everything going well on your end and with your family?