Learning About Outgroups:

The Impact of Broad Versus Deep Interactions*

Anujit Chakraborty† Arkadev Ghosh‡ Matt Lowe§ Gareth Nellis¶

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. While deep contact strengthens social and economic ties with the outgroup partner interacted with, only broad contact reduces misperceptions about outgroup strangers. These findings align with a model in which independent sampling (observing multiple outgroup members) promotes learning about outgroups more than prolonged interaction with a single individual does. Nevertheless, neither type of contact changes behavior toward the wider outgroup.

*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 Sid Baral and Karen Ferree.

†University of California, Davis. Email: chakraborty@ucdavis.edu.

‡Duke University. Email: arkadev.ghosh@duke.edu.

§University of British Columbia. Email: matt.lowe@ubc.ca.

¶University of California San Diego. Email: gnellis@ucsd.edu.

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. 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). The lack of knowledge that groups frequently have about one another, and the distorted imaginations and misrepresentations that result, threatens to skew intergroup relations, leading to suboptimal behaviors and decision-making. The consequences may be severe. 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).

Collaborative contact has emerged as a leading tool for reducing intergroup enmity (Allport, 1954; Mousa, 2020; Lowe, 2021). By fostering familiarity, 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). 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, 2024). 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 that begins with a simple observation. 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; Paluck and Clark, 2020), perpetuating inaccurate beliefs about that group. This echoes the trope, "I'm not racist because I have a Black friend." It stands to reason that positive 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, and note that it may operate through both Bayesian and non-Bayesian processes. 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 tack. 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, 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. In short, there is little consensus on the relative importance of the sampling versus friendship channels. Hence, there is uncertainty about which type of contact—broad or deep—is ultimately better for improving intergroup relations.

To formalize these ideas, we present a simple model of how contact shapes belief updating about outgroups. We distinguish between highly visible attributes of outgroups (e.g., competence traits like productivity in a work setting) versus characteristics that are less easily observed (e.g., moral traits like honesty and decency). Holding constant the total time spent interacting with outgroup individuals, the model shows three things. First, when it comes to updating about highly visible attributes, there is a *sampling trade-off* between broad contact—which offers a larger number of independent sample points—and deep contact, which offers greater precision for any given sample point. Broad contact thus leads to generalized learning about the wider outgroup for highly visible group characteristics, even as this shorter-duration contact limits learning about any particular outgroup individual interacted with. Second, the trade-off extends to the case of updating about friendship potential, a bundle of less visible characteristics. Learning about a contact partner's friendship potential demands costly investment by a decision-maker. Such investment is only worthwhile when any resulting friendship can pay off over time. This means that only sufficiently deep contact facilitates the richer conversations needed for friendship potential to be discovered—and for cross-group friendship to materialize. But third, the sampling trade-off implies that such learning will not generalize beyond specific contact partners. For that to happen, total contact hours would have to be expanded to enable an intervention that is both deep and broad concurrently.

We conducted a field experiment in India, among men between the ages of 18 and 40, to study the benefits and trade-offs of broad versus deep contact. Our study site is North 24 Parganas, a West Bengal district that lies on the India-Bangladesh border. It is a region where tensions between Hindus (the majority group) and Muslims (the minority group) run deep and have erupted into violence at several points over the past decade. We set up a factory in which participants worked on an intensive joint production task in pairs for six days. The task involved making paper bags, and required constant physical interaction and coordination between partners. For the main randomization, we assigned 964 subjects to either work with the same outgroup partner for six days, or with six different outgroup partners (a new partner each day). In a third arm, majority-group individuals (Hindus) worked for six days with the same ingroup partner. A fourth arm (comprising both Hindus and Muslims) served as a pure control condition. 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 separate "holdout" sample—as well as social and economic ties, willingness to interact with outgroup members, and altruism.

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* conditions. They have shallower conversations: they discuss fewer topics overall than their deep counterparts, and their conversations are much less likely to have covered topics like TV, movies, music, sports, hobbies, and politics. Perhaps as a result, broad subjects report being less happy (-0.09σ, $p = 0.06$) and less friendly with their work partners (-0.47σ, $p < 0.01$) 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. The lower camaraderie seen in the broad condition resonates with work in organizational behavior, which highlights associations between low staff morale, an absence of close friendships at work, and high employee turnover (e.g., Harter et al., 2002).

The differences in workplace atmosphere translate into different social and economic ties longer-term. At endline, 60% of *Outgroup, deep* subjects stated that they had 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 highly conducive to stimulating intergroup social and economic ties—a vital input into social cohesion—whereas broad contact falls flat, relatively speaking.

In contrast, only broad contact promotes learning about outgroup characteristics. 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* (0.09σ, $p = 0.03$). These effects are driven by guesses about outgroup attributes that were straightforward to observe in the factory setting: 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 Hindus in the holdout sample. This null finding makes sense. We show 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. In a placebo test, we find that neither of the outgroup conditions increases prediction accuracy about ingroups. In an encouraging result for intergroup contact in general, we find that subjects assigned to either of the outgroup conditions (*Outgroup, deep* or *Outgroup, broad*) report more similar guesses about ingroups and outgroups in the holdout sample. It appears that contact of any sort engenders recognition of a common humanity—countering the tendency to view outgroups as fundamentally alien from ingroups.

Neither type of contact changes broader intergroup behaviors: 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. Unexpectedly, we also estimate null effects of broad contact on productivity. Despite participants' more critical assessment of their own and their partner's production in broad relative to deep, factory managers' actual counts of bags produced, as well as their quality ratings, show no statistical differences between the two treatment arms.

Our overall message is clear: deep contact with outgroups is better for building intergroup ties, while only broad contact corrects misperceptions about the outgroup at large. Productivity is similar regardless of the type of contact. The optimal choice then depends on whether the policymaker's primary objective is to nurture intergroup connections or to counter wrong beliefs.

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. In a direct sense, they can give a fillip to miguided policies aimed at addressing the "problem" identified in a stereotype (for example, the belief that minorities have too many children might spur sterilization campaigns). Indirectly, rumors that malign certain groups can fuel a climate of hate, providing rationales and rhetorical cover for the mistreatment of targeted groups (cf. Bursztyn et al., 2023). More pessimistically, however, our results underscore the challenges to engineering generalized behavioral change via contact alone. At the heart of the matter is a Catch-22. Expansive sampling (through broad interaction) is informative about outgroups, but only for visible group traits; friendship (through deep interactions) offers richer connections with individual outgroup members, 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. Frontier research on intergroup contact explores the mediating role of three of the four necessary conditions for intergroup contact to reduce prejudice, as posited by Allport (1954).

Specifically, common goals (Lowe, 2021), intergroup cooperation (Ghosh, 2022), and equal status (Greene et al., 2023) have been shown to amplify contact's positive effects. Nevertheless, recent meta-analytic work finds that even contact that satsifies all four Allport conditions has limited generalized effects (Lowe, 2024), 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 generalization, whereas the friendship channel does not—countering prominent claims in psychology. Given the pervasiveness of intergroup contact in real-world settings like schools (Alan et al., 2021), worksites (Thachil, 2017), and militaries (Lyall, 2020), there are major policy gains to realizing how best to structure it.

Beyond beliefs about outgroups, many papers document misperceptions about others, 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. A notable example is Matavelli (2024), who finds that discussions with peers about masculinity quickly reduce misperceptions around peers' opinions on masculinity norms. We focus instead on the tradeoff between sampling and friendship potential when learning a broad set of outgroup characteristics.

Next, our study is guided by—and extends—theoretical work on stereotypes (Bordalo et al., 2016). Rational perspectives depict stereotypes as originating in statistical discrimination, with decision-makers applying group characteristics to individuals. Such models assume these beliefs are essentially correct, thus ruling out the possibility of inaccuracy or further learning about the aggregate distribution. However, recent research on inaccurate statistical discrimination shows that stereotypes can be flawed and subject to change (Bohren et al., 2019, 2023). Our conceptual framework aligns with this literature by modeling inaccurate stereotyping, allowing for stereotypes to evolve. We follow *social cognition* approaches that view stereotypes as heuristics that individuals deploy in their everyday lives to save on cognitive resources. Such stereotypes are based on group features that are the most distinctive and easiest to recall, and often entail exaggerations of true group differences (Judd and Park, 1993). Our broad treatment was designed to expose subjects to interactions with multiple outgroup members, to help them break out of such exaggerated thinking.

Finally, our experiment design also constitutes a novel test of the strength of weak ties hypothesis: the idea that mere acquaintances aid job transitions by providing diverse and novel information (Granovetter, 1973). Causal tests of the weak ties theory are rare, though a notable example is the LinkedIn experiment of Rajkumar et al. (2022), which finds that weak (online) ties create more job mobility than strong ties. While we find that weak ties (broad contact) lead to more learning about the outgroup than strong ties (deep contact), our closest parallel measure to

Rajkumar et al. (2022) leads to the opposite conclusion: participants are as, or more, likely to have job-finding help in the strong-tie condition than in the weak-tie condition.

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 (panchayats), 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).[1] Electoral politics, too, has caused social rifts.[2] Last, 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.[3]

2.2 Factory setting

For the intervention, we set up our own "factory in the field."[4] 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.

2.3 Production task

Workers made two types of paper bags, similar to the kind used to bag up groceries in local stores in the town where we worked (see Appendix Figure S9 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.

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

[2] 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.

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

[4] We decided to establish our own factory rather than partnering with an existing one because we required a high level of control over the intervention—this control would be near-impossible for an existing firm to accommodate by changing their standard operating practices.

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 rationale for having subjects produce small bags is that they provide a measure of productivity at the individual level. At the same time, we wanted 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, which is what the big bags encourage. Still, even for the small bags we took steps to encourage coordination and interaction. Within each booth, pairs of workers were 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.

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.

2.4 Worker compensation

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),[5] and a bonus of Rs. 200 for workers who attended for all nine days of work. The state minimum daily wage for unskilled labor at the time of the experiment was Rs. 376, although in reality very few employers abide by this wage. Also, such a daily wage would typically involve at least eight hours of work. Thus, given both the good working conditions and high pay, this was seen as a desirable job by most people we approached with an offer in Barasat.

2.5 Recruitment and screening

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 to express their interest. Upon completing the second baseline survey, they were

[5] 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.

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. These two stages of selection select participants that are somewhat more educated and less likely to be employed, though there is no evidence of selection along dimensions related to prejudice or political views (Appendix Table S1).

2.6 Implementation

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). Additionally, they 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.[6]

2.7 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

[6]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.

informed on the very first day that they 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.[7]

4. *Pure control.* After the three-day lead-in period, these participants were randomized to serve as the pure control group. In other words, they did not participate in factory work during the following six days, unlike the other three treatment groups.

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

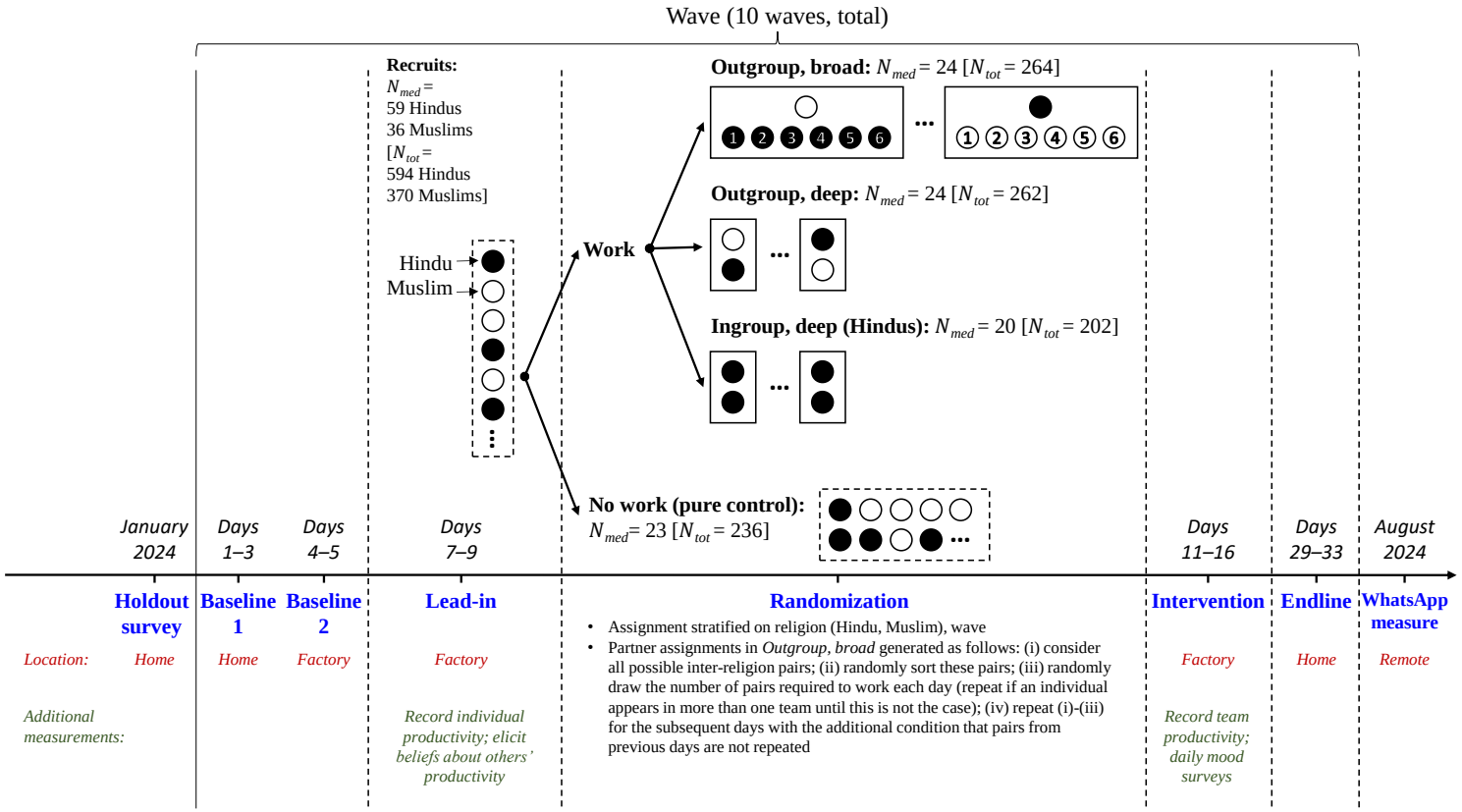
1. Consider all possible inter-religious pairs in the *Outgroup, broad* arm.
2. Randomly sort these pairs.
3. Randomly draw the number of pairs required to work each day; repeat if an individual appears in more than one pair until this is not the case.
4. Follow steps 1 to 3 above for the next days with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

Two 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 below.

[7] 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).

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*.

3 Data and empirical strategy

3.1 Outcomes

Daily mood surveys During the six-day intervention period, workers completed daily 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 S4). 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 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, we presented workers with five pairs of names (four mixed-religion pairs[8] 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.

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. Again, we asked broad participants only about their partners from days one, four, and

[8]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.

six.

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 Barasat 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 S9 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.[9] We carefully describe 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).[10]

We collected nine measures of misperceptions across seven domains (Table 1). 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 Figure 2, we show that, in reality, Hindus and Muslims behave similarly. The only significant differences we see are that Muslims cheat less than Hindus (reporting 5.72 versus 6.46 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

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

[10]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 1: 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 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)
Risk preferences	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)
	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)
Generosity	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)

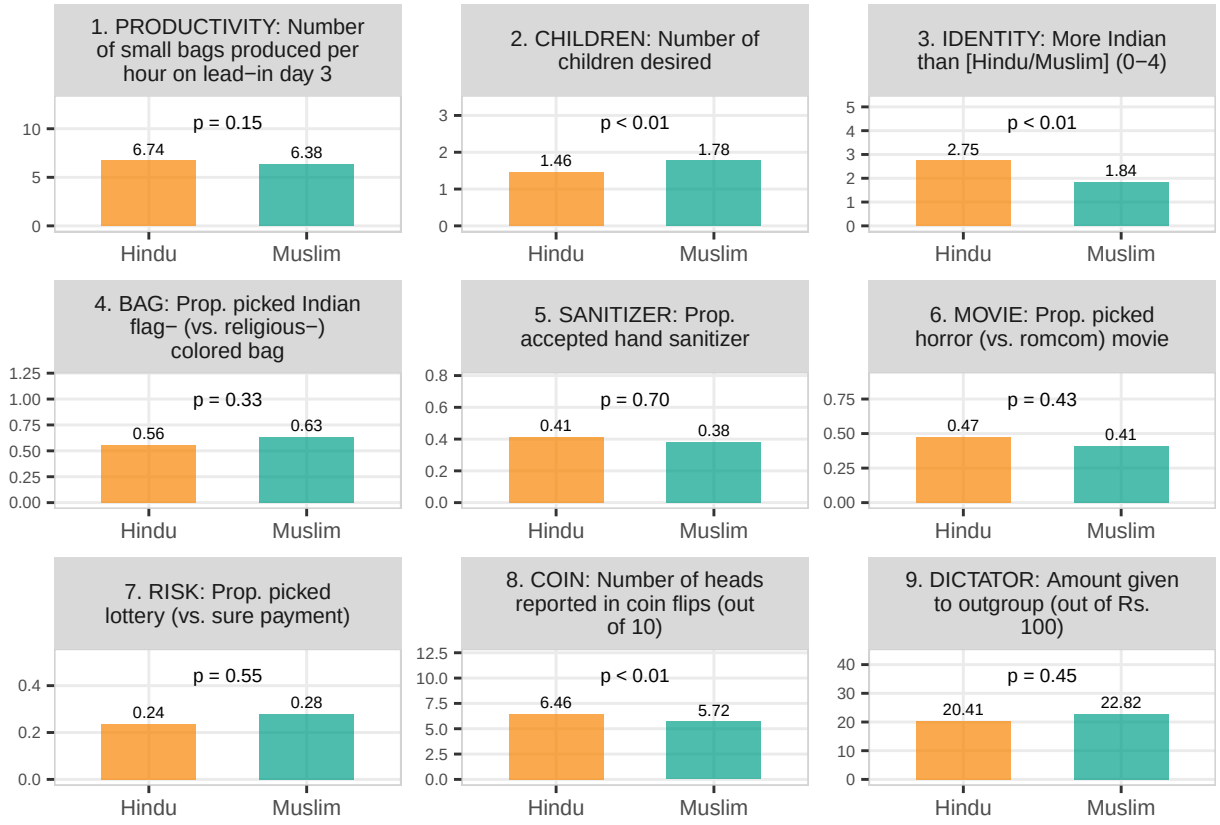
that stereotypical beliefs can contain a kernel of truth (Bordalo et al. 2016).

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 of Table 1.[11] 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: cheating, risk preferences, and altruism. We classify more-visible attributes as those that can be readily

[11] 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 S8, 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).

Figure 2: Preferences and beliefs 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.

inferred at work: productivity (since this is the main factory activity), 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). We also consider fertility preferences to be a more visible attribute since family is a ubiquitous conversation topic in Bengal.[12]

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

[12]We validate this in Appendix Figure S5, which shows that family relations were the second most talked about topic between factory partners, after "work."

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 agreed to a price, we stopped asking further. We define the reservation wage as the midpoint between the two wage offers that causes 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.6), 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. 51 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).

India's Hindu nationalist Prime Minister, Narendra Modi, is viewed favorably by most study participants, with an average feeling thermometer score of 67 out of 100. But there is a large difference in attitudes between sampled Hindus (average Modi rating of 85) and Muslims (average of 39). Moreover, Appendix Figure S3 shows evidence of substantial outgroup bias, as well as asymmetry in the magnitude of that bias. When asked to rank the trustworthiness of five groups (with one as the highest rank), Muslims rank Muslims as 1.1 on average and Hindus as 2.5 (a difference of 1.4). Hindus, meanwhile, rank Hindus as 1.1 on average and Muslims as 3.4 (a difference of 2.3). Thus, the majority distrusts the minority more than vice versa.

Appendix Table S3 presents an analysis of balance on pre-treatment covariates. We fail to reject the null hypothesis—that no experimental arm has a significant effect on a given covariate at the 0.05 alpha level—for

all 19 of the baseline covariates analyzed, consistent with random assignment having formed similar groups.

Endline attrition was very low, with 95% of subjects completing the endline survey. Appendix Table S4 shows that attrition is uncorrelated with treatment assignment, and that attrition did not measurably differ between Hindus and Muslims.

3.3 Compliance

Our efforts to keep daily attendance at the factory high—by screening for high-attendance types using the three lead-in days, paying good salaries, and fielding regular follow-up calls to those who did not show up to work—were successful. Overall attendance averaged 96%. In Appendix Table S4, we find no signs of differential attendance between Hindus and Muslims. However, attendance was marginally significantly higher in *Outgroup, broad* than in *Outgroup, deep* ($p = 0.06$ for the difference). It may be that, for a few individuals, spending six days with an outgroup partner they do not get along with (as in deep) is harder than spending just one day with an unappealing partner (as in broad). Note, we estimate intent-to-treat effects, and this small difference has no bearing on the internal validity of our estimates.

3.4 Pre-registration

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. Minor deviations from the pre-registration are detailed in Appendix C.1.

3.5 Empirical specification

We examine the effects of broad versus deep contact on attitudes, beliefs, and behaviors at endline using the following specification:

$$y_i = \psi_s + \beta_{ID} Ingroup,\ deep_i + \beta_{OD} Outgroup,\ deep_i + \beta_{OB} Outgroup,\ broad_i + \varepsilon_i \quad (1)$$

where y_i denotes the outcome recorded for individual i, and ψ_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. β_{ID} is identified only by Hindu participants, since no Muslims are assigned to the *Ingroup, deep* condition. In some analyses, we restrict the sample only to individuals assigned to the *Outgroup, deep* and *Outgroup, broad* conditions, for simplicity of presentation. We report robust standard errors.

For analyses employing the daily mood surveys, we report coefficients from this specification:

$$y_{i,t} = \mu_s + \gamma_{OD} Outgroup,\ deep_i + \gamma_{OB} Outgroup,\ broad_i + u_{i,t} \tag{2}$$

Outcomes are measured daily, with $t \in \{1,2,3,4,5,6\}$, and *Ingroup, deep* serves as the omitted category. The strata (μ_s) are religion $\times$ wave $\times$ day $\times$ shift fixed effects (recall that the religion $\times$ wave fixed effects are the randomization strata). Standard errors are clustered by individual. We drop observations for individuals whose partner was absent on that day.

4 Theory

In this section, we consider a decision-maker (DM) who is estimating the mean value for some characteristic of outgroup individuals. Subsection 4.1 shows that when it comes to updating about *visible attributes*, deeper versus broader interactions offer an intuitive tradeoff: broader interactions facilitate generalized updating about the outgroup at large, but the shortness of the interactions restricts the updating possible about each matched partner. Deeper interactions facilitate updating about each matched partner but the smaller sample size of partners restricts the scope for updating about the larger outgroup population. Subsection 4.2 extends this discussion to the case of updating about friendship potential, which is a non-visible attribute, and hence the DM needs to pay a cost of interaction to learn about every partner's friendship potential. Paying such a cost is only worthwhile when the friendship pays off over a longer period of interaction. Thus, when it comes to non-visible attributes, only sufficiently deep interactions facilitate learning about the partner, but the requirement of deep interactions automatically limits the number of distinct outgroup partners one can interact with and hence limits generalized learning about the outgroup.

4.1 Visible attributes

The DM starts with an inaccurate stereotyped estimate $x_0^{st} \neq x^*$ about the outgroup initially, and (falsely) believes that $x_0^{st} = x^* + e^{st}$, where e^{st} is mean zero noise. We assume that the DM interacts with $n \geq 1$ outgroup members, and meets with each outgroup member $D \geq 1$ times. n reflects the breadth of outgroup contact, while D reflects the depth. The DM thus has $K = nD$ total interactions with outgroup members.

Noisy observations If the characteristic is *visible* during interactions, then the DM takes account of those observations to gradually learn the average characteristic. 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. While the mean

characteristic of all outgroup members is x^*, 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 η_i and ε_{id} are individual- and occasion-specific noise terms. We will write $x^* + \eta_m$ as the true characteristic of individual m, a generic outgroup member that the DM never encounters.

DM's model of the world Let E denote the expectation with respect to the joint distribution of all e, ε, η variables, for *all* individuals including m. We will use ε and η as a short hand for the collection of all random variables $(\varepsilon_{id})_{i,d}$ and $(\eta_i)_i$ respectively. The DM makes the following assumptions about the world: $E(e^{st}|\varepsilon,\eta) = 0$, $E((e^{st})^2|\varepsilon,\eta) = \sigma_{st}^2$. Next, $E(\eta_i|e^{st},\varepsilon) = 0$, $E(\eta_i^2|e^{st},\varepsilon) = \sigma_{het}^2$, $E(\eta_i\eta_j|e^{st},\varepsilon) = 0$. Finally, $E(\varepsilon_{id}|e^{st},\eta) = 0$, $E(\varepsilon_{id}^2|e^{st},\eta) = \sigma_{inf}^2$, and unless $i_1 = i_2$ and $d_1 = d_2$, $E(\varepsilon_{i_1 d_1}\varepsilon_{i_2 d_2}|e^{st},\eta) = 0$. The conditional means being zero imply that the unconditional means are also zero. σ_{inf}^2 is a measure of how noisy the observations are and σ_{het}^2 is a measure of how heterogeneous the outgroup characteristic is. The DM uses these assumptions to form an estimate $\widehat{x_{n,D}}$ of the attribute for outgroup individual i. We define a notion of accuracy that benchmarks every estimate against the truth. Thus, both bias (uni-directional mistakes) and noise (general mistakes) reduce accuracy.

Definition 1. *The stereotype-affected accuracy when estimating the attribute for person i is defined as the negative of the variance, that is* $\mathscr{A}(\widehat{x_{n,D}}) = -\mathscr{V}(\widehat{x_{n,D}}) = -E(\widehat{x_{n,D}} - x^* - \eta_i)^2$.

DM's Estimate for any n, D To update her estimate 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 minimize the variance $\mathscr{V}$ of the weighted "unbiased"[13] estimator

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

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

Minimizing variance requires that the DM's estimate about the generic outgroup member m gets updated as

$$\begin{aligned}
\widehat{x_{n,D}}^{gen} &= \arg\max_{(w^{st},(w_{id})_{i,d})} -E\left(\widehat{x}(w^{st},(w_{id})_{i,d}) - x^* - \eta_m\right)^2 \\
&= \arg\max_{(w^{st},(w_{id})_{i,d})} -E\left(w^{st}(x_0^{st}-x^*) + \sum_{i\leq n}\sum_{d\leq D} w_{id}(x_{id}-x^*) - \eta_m\right)^2 \\
&= \arg\max_{(w^{st},(w_{id})_{i,d})} -E\left(w^{st}(x_0^{st}-x^*) + \sum_{i\leq n}\sum_{d\leq D} w_{id}(x_{id}-x^*)\right)^2 \\
&= \arg\max_{(w^{st},(w_{id})_{i,d})} -E\left(w^{st}(\varepsilon_0^{st}) + \sum_{i\leq n}\sum_{d\leq D} w_{id}(\eta_i+\varepsilon_{id})\right)^2 \\
&= \arg\max_{(w^{st},(w_{id})_{i,d})} -\left((w^{st})^2E(\varepsilon_0^{st})^2 + \sum_{i\leq n}\sum_{d\leq D}(w_{id})^2E(\eta_i^2+\varepsilon_{id}^2) + \sum_{i\leq n}\sum_{d\neq d'}(2w_{id}w_{id'})E(\eta_i^2)\right) \qquad (3) \\
&= \frac{1}{\frac{1}{\sigma_{st}^2} + \frac{n}{\sigma_{inf}^2/D+\sigma_{het}^2}}\left(\frac{1}{\sigma_{st}^2}\times x_0^{st} + \sum_{i\leq n}\frac{1}{\sigma_{inf}^2/D+\sigma_{het}^2}\times\frac{\sum_{d\leq D}x_{id}}{D}\right) \qquad (4)
\end{aligned}$$

Appendix C.2 contains the complete derivation. To give the reader some idea of how the proof works, we only show a rough outline for the steps leading to equation (3). In particular, because η_m is uncorrelated with everything piece of information, it becomes irrelevant to the optimal weighting rule. After equation (3), one needs to calculate the $(nD+1)$ derivatives of equation (3) with respect to the $(nD+1)$ weights $w^{st},(w_{id})_{i,d}$, and set them all equal to zero. As the system of $(nD+1)$ derivatives obtained are invariant to interchanging any pair of coefficients $w_{id},w_{i'd'}$, the solution to that system of equations features $w_{id}=w_{i'd'}$ for any two weights.

Overall, as shown in the final equation (4), maximizing accuracy $\mathscr{A}$ requires weighting each *independent* piece of information, be it the initial measure x_{st}, or each of the n sample averages, $\frac{\sum_{d\leq D}x_{id}}{D}$, by the inverse of its corresponding variance σ_{st}^2 and $\sigma_{inf}^2/D+\sigma_{het}^2$ respectively. The whole expression is then normalized so that the total weights add up to 1. With such inverse variance weights, more precise information gets more importance.[14] For example, when the prior is considered to be noisier (higher σ_{st}), or when the observations are more precise (lower σ_{inf}), or when the population is less heterogeneous (lower σ_{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 estimate about *any one* of the n

[14]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.

individuals she has interacted with (say individual I):

$$\begin{aligned}\widehat{x_{n,D}}^{I} &= \arg\max_{w^{st},(w_{id})_{i,d}} -E\left(\widehat{x}(w^{st},(w_{id})_{i,d}) - (x^{*}+\eta_{I})\right)^{2} \\ &= \arg\max_{w^{st},(w_{id})_{i,d}} -E\left(w^{st}(x_{0}^{st}-x^{*}-\eta_{I}) + \sum_{i\leq n}\sum_{d\leq D} w_{id}(x_{id}-x^{*}-\eta_{I})\right)^{2} \\ &= \arg\max_{w^{st},(w_{id})_{i,d}} -E\left(w^{st}(\varepsilon_{0}^{st}-\eta_{I}) + \sum_{i\neq I}\sum_{d\leq D} w_{id}(\eta_{i}+\varepsilon_{id}-\eta_{I}) + \sum_{d\leq D} w_{Id}\varepsilon_{Id}\right)^{2}\end{aligned}$$

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

$$w_{id} = \frac{\sigma_{st}^{2}\sigma_{inf}^{2}}{(\sigma_{inf}^{2}+D\sigma_{het}^{2})(nD\sigma_{st}^{2}+\sigma_{inf}^{2}+D\sigma_{het}^{2})},\ \ w_{Id} = \frac{\sigma_{st}^{2}\sigma_{inf}^{2}+nD\sigma_{st}^{2}\sigma_{het}^{2}+\sigma_{inf}^{2}\sigma_{het}^{2}+D\sigma_{het}^{4}}{(\sigma_{inf}^{2}+D\sigma_{het}^{2})(nD\sigma_{st}^{2}+\sigma_{inf}^{2}+D\sigma_{het}^{2})},\ \ w^{st} = \frac{\sigma_{inf}^{2}}{nD\sigma_{st}^{2}+\sigma_{inf}^{2}+D\sigma_{het}^{2}}$$

Appendix C.3 contains the complete derivation. Even though the DM has encountered n individuals in both cases, the estimate $\widehat{x_{n,D}}^{I}$ about an encountered person is different from the estimate about the general outgroup $\widehat{x_{n,D}}^{gen}$. In particular, $\sigma_{het} > 0$ implies $w_{Id} > w_{id}$ for $i \neq I$. That is, all x_{Id} is informationally favored over all x_{id}, as only the former informs about η_{I}. Relatedly, the error/noise in x_{0}^{st} is now correlated to that in the sample average obtained from any non-I outgroup member, as both fail to account for η_{I}.[15] The mathematical steps are identical to the last case: calculate accuracy, calculate its derivatives with respect to the weights, and then solve the system of derived equations.

Extent of generalization after n interactions Before the n interactions took place, the person I to be encountered was just another random outgroup person, for whom the DM held an estimate of x^{st}. After meeting n individuals, the DM updates both about that individual I, as well as the outgroup in general. After that cumulative experience, how does the extent of updating, measured in terms of accuracy, over an encountered individual generalize to the updating over a random person from the entire outgroup? We define the *generalizability factor* after n interactions by the relative quantities of accuracy in $\mathscr{A}^{I}$ as compared to that in $\mathscr{A}^{gen}$, after n meetings: $\mathscr{G}(n,D) = \frac{\mathscr{V}^{I}}{\mathscr{V}^{gen}}$. For a fixed $\mathscr{V}^{I}$, a higher $\mathscr{V}^{gen}$ implies a smaller generalization. Further, this fraction *always* lies between 0 and 1, and is equal to 1 only when $\sigma_{het} = 0$, that is, when all individuals are identical and hence learning about one is equivalent to learning about all.

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, broader contact (increasing n) or deeper contact (increasing D), both

[15] For comparison, when updating about the stranger m, all information sources were equally (un)-informative about η_{m}.

increase the accuracy about the outgroup at large. Further, increasing n has diminishing returns, i.e, on the margin, the accuracy-increasing effect of contact is smaller for individuals who have already encountered a higher number of outgroup members. In our setting, this 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 reduce the bias about an encountered outgroup member (subject to the outgroup heterogeneity σ_{het} being large enough). Thus, there is a trade-off between how much the DM can update about each encountered person, and how much they can generalize that to the wider group.

Result 1. *Fix x_{st} and $b = |x_{st} - x^*|$. Let $\mathscr{A}^{gen}$ and $\mathscr{A}^I$ be the optimal accuracy when updating about m and I respectively. For visible attributes, the following results hold:*
i) Returns to contact: For any fixed D and for $b^2 \geq \frac{2}{3}\sigma_{st}^2$, increasing n increases $\mathscr{A}^{gen}$ but with diminishing returns. For any fixed n and for $b^2 \geq \frac{2}{3}\sigma_{st}^2$, increasing D also increases $\mathscr{A}^{gen}$.
ii) n versus D Tradeoff in $\mathscr{A}^{gen}$: Fix $K = nD$. $b^2 \geq \frac{2}{3}\sigma_{st}^2$ guarantees that increasing n increases $\mathscr{A}^{gen}$ iii) n versus D Tradeoff in $\mathscr{B}^I$ and $\mathscr{G}(n,D)$: Fix $K = nD$. $\sigma_{het}^2 \geq \sigma_{inf}^2$ guarantees that $\mathscr{A}^I$ increases as D increases. Therefore, under $b^2 \geq \frac{2}{3}\sigma_{st}^2$ and $\sigma_{het}^2 \geq \sigma_{inf}^2$, the generalizability factor $\mathscr{G}(n,D) = \frac{\mathscr{V}^I}{\mathscr{V}^{gen}}$ decreases as D increases.

Proof. See Appendix C.4. □

Discussion In the model misperceptions originate in noisy observations and the prior (stereotypical) parameter x_0^{st} being different from x^*, while the process of updating the estimate 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.

4.2 Friendship potential as a non-visible attribute

Each interaction with an outgroup member not only provides information about visible characteristics, but also allows an opportunity to form a new friendship. We assume that friendship formation requires investing time and (cognitive) effort in seeking commonalities with the other person. While the investment is costly, when successful in forming a friendship, it increases the utility of the DM throughout the period of the interaction. 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 cost of investment in friendship, the quality of friendship formed or friendship potential, and the duration of the relationship through which the (relative) benefits can be enjoyed.

As in the discussion in Subsection 4.1 about *visible attributes*, we assume that the true friendship potential in the outgroup is x^*, but the DM starts with a stereotypically inaccurate estimate x_f^{st}. We also assume the same model of the world and method of updating, with one crucial difference: Friendship potential is not physically visible from one's appearance, and hence can only be observed and then used for updating when the DM invests into the friendship.

Investment in friendship At the beginning of any interaction with outgroup member i, the DM makes a myopic binary choice between investing in a potential friendship or not, and chooses to invest if $\hat{x}^i D - c \geq 0$, or $D \geq c/\hat{x}^i = \bar{D}$, where $\hat{x}_f^{i-1}$ is her (updated) estimate of friendship potential at the beginning of the i-th interaction, $\hat{x}^i D$ is the cumulative utility from all interactions, and c is the cost of investment. Simply put, investment is only worth it if the duration of interaction D is deep enough.

Bad belief trap The requirement for deep enough contact to induce learning about friendship potential, along with our Result 1 about updating, creates a "bad belief trap" when it comes to the scope of long-term interactions with the outgroup. First, subjects who start with a stereotypically low x_f^{st} estimate of friendship potential, are less likely to find investing in friendships worthwhile, and hence get no opportunities to update. Thus, beginning with an incorrectly low opinion of friendship potential can be a steady state in itself, trapping DMs in that wrong belief and stops them from seeking more information. Second, only interactions that are deep enough, $D \geq \bar{D}$, are found to be worth investing into. But as discussed in Result 1, deep interactions limit generalizability. Thus, even a one-time deep intervention (for example, our *Outgroup, deep* treatment) that utilizes $D \geq \bar{D}$ to induce the DM to befriend that partner, fails to create far-reaching effects that generalize towards the outgroup, as the DM updates little about a random outgroup member from that interaction. The only way the DM could update positively about the friendship potential of the whole outgroup, is if she is exposed to an intervention that is both sufficiently deep and sufficiently

broad.

Predictions about friendship This simple model leads to the following predictions: subjects in the *Outgroup, deep* treatment should be more likely to invest in friendship than those in *Outgroup, broad*, for example, by engaging in a higher number of meaningful conversations, and this would lead to closer ties, a happier work environment, and higher future interactions with matched partners outside of the factory context. Still, we predict that this attitude change towards an outgroup partner will not generalize beyond the matched partner towards the larger outgroup under the *Outgroup, deep* intervention of $n = 1$.

5 Empirical results

5.1 Daily mood

We first analyze impacts on participants' subjective assessments of their mood and work performance. At the end of each work day, we administered a brief survey in which workers privately assessed their mood and rated their own and their partner's performance on that day's task. For all measures, the estimates suggest that assignment to the *Outgroup, broad* condition produced an environment and a set of emotions that were substantially different from those 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 (compared to deep) condition (columns 4 and 5, both $p < 0.01$).

Deep and broad contact differ in 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. 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 0.16σ happier ($p = 0.09$), and report their partner to be 0.17σ closer as a friend ($p < 0.05$). Merely anticipating future outgroup interaction thus meaningfully affects current mood and relationships.

Our theory predicts that deep contact should lay the foundations for partner discussions that are more intimate,

more fun, and more friendship-inducing; when you know that you have to spend a longer time to spend with someone, you are more willing to invest in them, and steer conversations away from platitudes. That is what we find. Table 2 examines treatment effects on the content of conversations with partners on day six, the final day of work. 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, $p = 0.01$), 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 stick to matters of work and personal life—talking about these topics at statistically similar rates to *Outgroup, deep* subjects (columns 3 and 4).

Table 2: *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 6 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.

Taken together, we find that the constant rotation of work partners gives rise to an atmosphere that is more isolating and less happy than a situation in which workers—whether belonging to an ingroup or outgroup—are permanently paired. However, 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. One possible explanation for this finding is that the anticipation of future interaction mitigates potential frictions caused by religious differences.

Table 3: Subjects' mood is 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 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 religion $\times$ wave $\times$ day $\times$ shift fixed effects. Robust standard errors, clustered by subject, are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

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. First, Hindus in the deep contact condition build similar ties with ingroup and outgroup partners. At endline, 66% of *Ingroup, deep* participants report having their partner's phone number, with no statistically significant difference relative to *Outgroup, deep* participants (Table 4, column 1). 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).

Table 4: *Outgroup, broad* reduces in-person and remote contact with work partners, and job and financial assistance.

	Outcome:				
	Has partner phone number (1)	Phone contact (2)	Food, family, in-person contact (3)	Work help (4)	Money help (5)
Outgroup, deep	-0.02	-0.01	-0.06	0.05*	0.02
	(0.05)	(0.05)	(0.04)	(0.03)	(0.02)
Outgroup, broad	-0.47***	-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	0.66	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

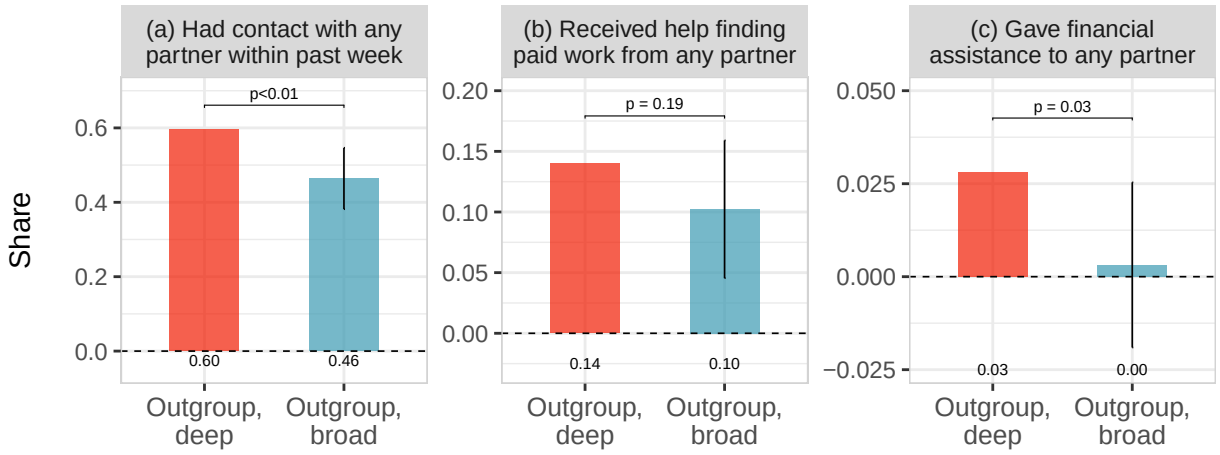
Notes: This table presents treatment effect estimates on recent interactions with work partners. The unit of analysis is the subject/partner pair. The survey question for columns 1–2 was: Please tell me which of the following types of contact (if any) you have had with [partner name] over the past week. 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 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 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 that the partner helped them to find paid work in any way since they worked together at the factory. 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. Regressions include randomization strata fixed effects. Robust standard errors, clustered by subject, are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

While having an outgroup partner does not hurt subsequent social ties, interacting only for one day rather than six does. *Outgroup, broad* participants are 47 percentage points less likely to have a given partner's phone number (they are asked about their partners on day 1, 4, and 6) than *Outgroup, deep* participants ($p < 0.01$, Table 4, column 1). Broad participants are 40 percentage points less likely to have had recent phone contact with a given partner, and 16 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 through their partners, and are less likely to borrow from or lend to a given partner than *Outgroup, deep* participants (columns 4 and 5).

Table 4 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 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

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



Notes: This figure shows estimated treatment effects on recent contact with former work partners. 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).

are 14 percentage points more likely to have had contact with any work partner over the last week ($p < 0.01$, Figure 3, panel (a)). *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)).[16]

The findings in Figure 3 speak against one particular version of the strength of weak ties theory (Granovetter 1973). In our setting, one strong outgroup tie dominates six weak outgroup ties. However, the findings in Table 4 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.

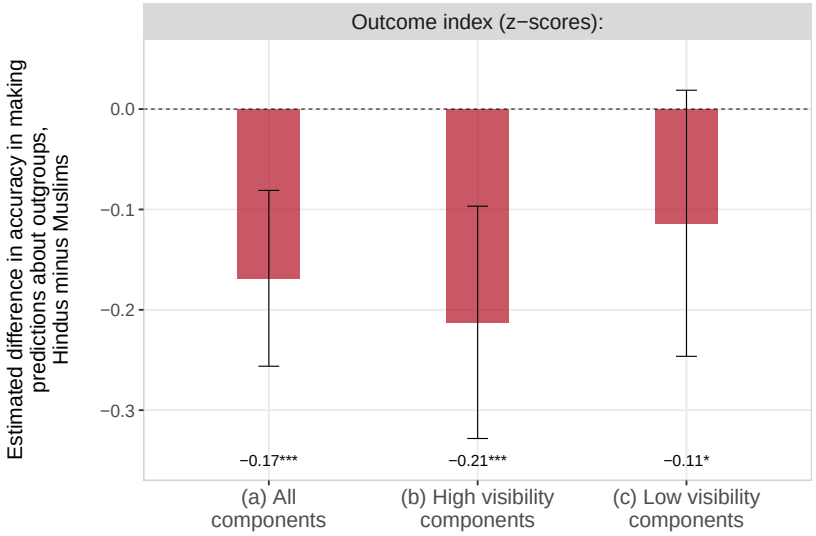
[16] Appendix Table S7 shows that there is no consistent pattern of heterogeneity in these effects by religion.

5.3 Beliefs about outgroups

Our results so far show that (i) participants build similar ties with fixed ingroup and outgroup partners, and (ii) deep contact leads to more positive work experiences, and stronger subsequent ties with partners than broad contact. We now turn to the question of whether broad contact leads to greater generalized belief updating. Here, again, we are guided by our model, which posits that there is a trade-off between deep and broad contact when it comes to updating about easily observable traits versus learning about a partner's friendship potential. Deep contact is more effective at creating stronger ties with partners, but does broad contact have stronger effects on generalized belief updating?

Before turning to the 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 exposed to many more Hindus—the majority—at work and in their residential wards than vice versa (Figure S1), meaning they have less scope for updating. This is also consistent with our theoretical framework.

Figure 4: 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 panel (a), all eight components on which subjects made guesses about the holdout sample, as well as their guesses about the productivity of other workers at the factory, are standardized and averaged in a single index. Panel (b) presents the mean of five specific components: productivity, bag, identity, children, and sanitizer. Panel (c) shows the mean of the remaining four components: coin, risk, movie, and dictator. *p<0.1; **p<0.05; ***p<0.01.

Figure 5 reports treatment effects on generalized belief updating. We show effects for *Outgroup, broad* versus *Outgroup, deep*—the two treatment conditions that involve outgroup exposure. The evidence strongly aligns with

the hypothesis of broad contact being more effective for generalized belief updating. Panel (i) shows that subjects assigned to *Outgroup, broad* are 0.08σ more accurate in their predictions about more visible attributes than those in *Outgroup, deep* ($p = 0.04$). There is no effect for the less visible attributes. Panels (ii) and (iii) show heterogeneous effects by religion. We observe a significant effect for Hindus, 0.21σ ($p < 0.01$), while the effect for Muslims is a precise null, aligning with the latter's higher accuracy about outgroup characteristics in the control group. In contrast, we estimate null effects of *Outgroup, deep* contact relative to the pure control group (Appendix Figure S6).[17] Only broad contact reduces misperceptions about the outgroup.

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 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. We estimate precise null effects for the full sample (panel (i), Appendix Figure S7) as well as for the heterogeneous effects by religion (panels (ii) and (iii)).

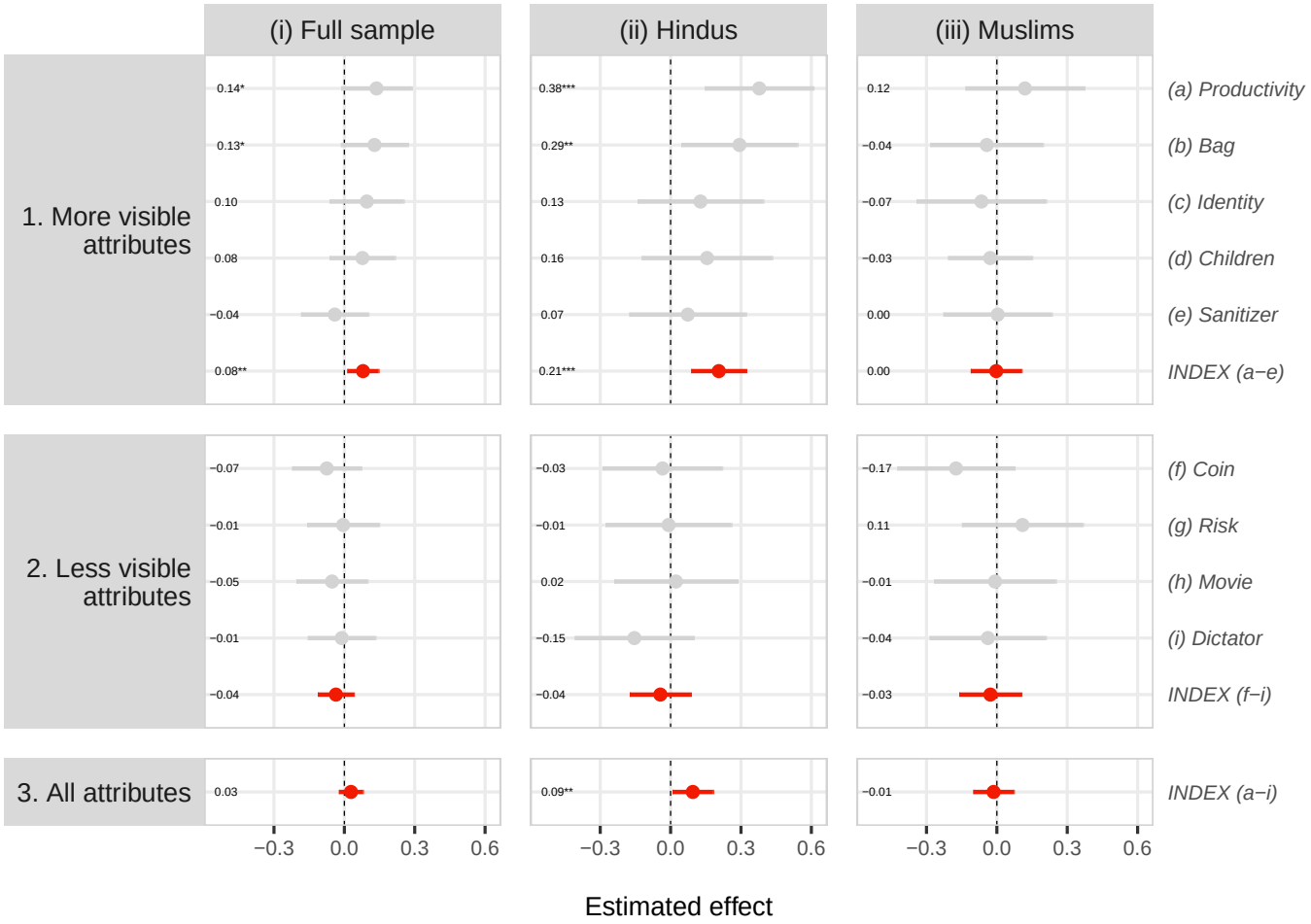
To change one's beliefs about outgroup strangers, one might expect it to be necessary to first alter beliefs about the individuals one has direct contact with. The *Outgroup, broad* treatment led Hindus to have more accurate beliefs about Muslim strangers, but did it also make them more accurate in their beliefs about their Muslim work partners? Further, is it the case that subjects in *Outgroup, deep* can predict more accurately about their work partners than subjects in *Outgroup, broad* (given the much longer exposure that deep entails)?

To test this, we asked treated participants to predict the *lead-in period* productivity of their partner from the six-day intervention period. For 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. Table 5 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. The fact that the effect sizes are very similar further suggests that workers were able to update about their partners fairly quickly—at least for the most visible characteristic in this experiment, productivity.

Our findings thus confirm that both types of contact facilitate belief updating about the individual interacted with. However, while participants in *Outgroup, broad* generalize their predictions to the broader group, those in the *Outgroup, deep* treatment do not. This suggests that for generalized belief updating, exposure to multiple out-

[17] 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 S6 is fully comparable with Figure 5.

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 is 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.

group members——emphasizing the importance of independent sampling——is more effective than exposure fostering deeper friendships.

Finally, 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

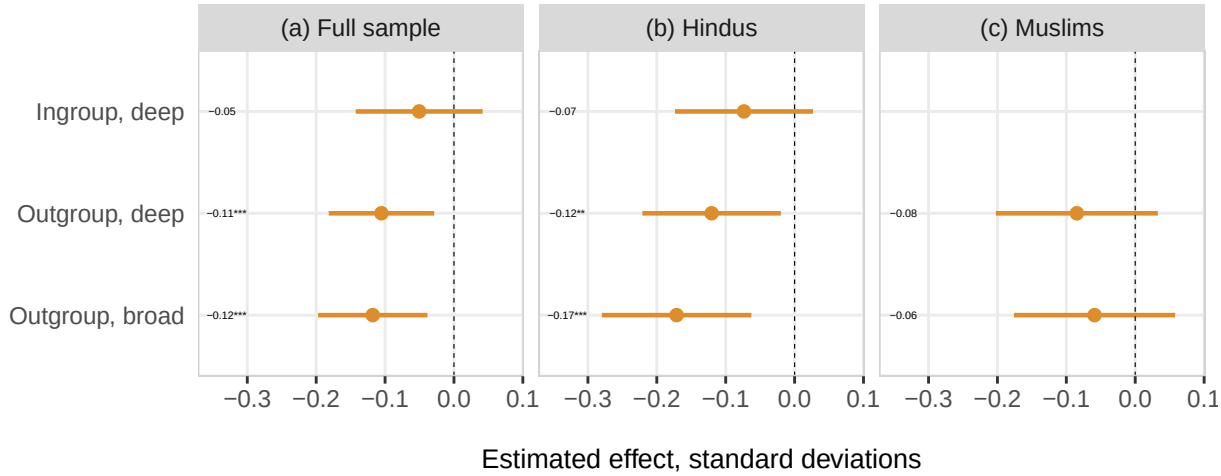
Table 5: *Outgroup, deep* and *Outgroup, broad* both increase prediction accuracy about partner productivity.

	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 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.

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, who have less baseline exposure to Muslims.

Figure 6: Both *deep* and *broad* contact with outgroups leads 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. The unit of analysis is the subject. 95% Confidence intervals are based on robust standard errors. *p<0.1; **p<0.05; ***p<0.01.

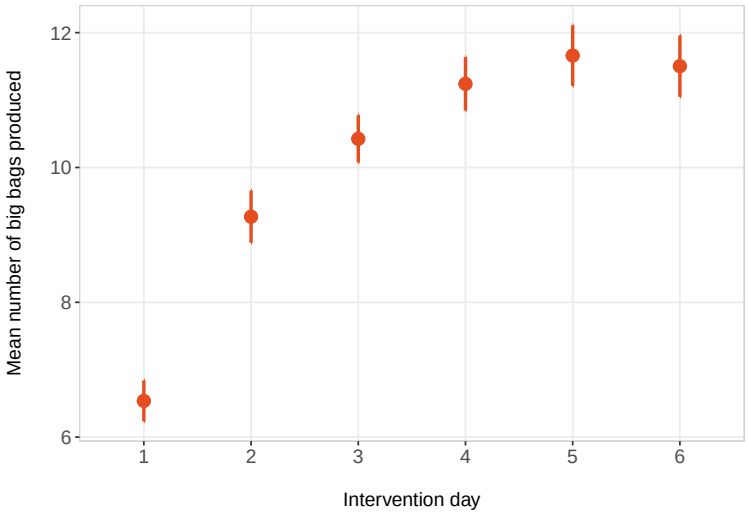
5.4 Does generalized belief updating come at the cost of lower productivity?

Religious diversity can negatively impact team production for several reasons. Workers may dislike working with non-coreligionists and seek to punish them, although fixed wages limit their ability to do so directly in this setting (Hjort, 2014). They may also be less concerned about their social reputation around outgroup members, which can result in reduced effort (Afridi et al., 2020). Additionally, workers might face higher coordination costs when collaborating with outgroup members or underestimate outgroup productivity, leading to lower effort provision when there are complementarities in production (Ghosh, 2022).[18] Furthermore, one may expect the negative effects of religious diversity to be more pronounced in *Outgroup, broad* compared to *Outgroup, deep*. This is because there is limited opportunity to form longer-term relationships with outgroup partners, making it harder to reduce these frictions over time.

Figure 7 shows that the average number of 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 type of Hindu-Muslim mixing impacts team productivity, using a regression in which we control for individual-level lead-in period productivity (Table 6). 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 and easy belief updating regarding the partner's productivity shortly after the start of the day's work. Indeed, our findings in Table 5 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.

[18] 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 S8).

Figure 7: 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 big bags produced on indicator variables for intervention day and demeaned randomization strata fixed effects. The unit of analysis is the worker-pair. 95% confidence intervals are based on standard errors clustered by unique worker-pair.

Table 6: Neither type of contact affects productivity.

	Outcome:		
	Number of big bags (1)	Big bag quality (2)	Number of big bags × quality (3)
Outgroup, deep	0.44	-0.06	2.58
	(0.40)	(0.07)	(3.35)
Outgroup, broad	0.28	-0.04	2.12
	(0.38)	(0.07)	(3.19)
N	1,989	1,987	1,987
Ingroup, deep mean	10.05	7.94	80.54
Outgroup, deep vs. Outgroup, broad *p*-value	0.63	0.71	0.87

Notes: This table reports the effects of treatments on production. We show only the effects on big bags—the effects on production of small bags are mechanically related given the requirement for pairs to produce two small bags for each big bag they made. Column 1 controls for the average number of bags produced by both partners during the lead-in period. Column 2 controls for the average quality of bags produced by both partners during the lead-in period. Column 3 controls 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. *p<0.1; **p<0.05; ***p<0.01.

5.5 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, 2024), and as predicted by our theoretical framework? We find evidence more consistent with the latter account in Table 7.

First, although we have shown that roughly one in two participants remain in contact with outgroup partners (Figure 3, panel (a)), this does not lead to additional close friendships with outgroup members, whether the contact was broad or deep (Table 7, column 1). This null result on close friendships is in stark contrast to the large, positive, and enduring, effects of youth camps on Hindu-Muslim friendships in the same municipality (Ghosh et al., 2023). 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. (2023) it was 15), or that the work environment is less conducive to friendship formation than the recreational camp environment.[19]

Second, we do not see evidence of either type of intergroup contact leading to an increase in generalized willingness to interact, as measured by incentivized willingness to work with an outgroup stranger (Table 7, column 2), or in generalized 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 0.24σ to 0.33σ 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.

To form more outgroup friendships or increase the willingness to work with outgroup strangers, individuals need a general sense of the friendship potential of those from the outgroup. Our theoretical model suggests that decision-makers are unlikely to generalize from just a few deep interactions to the group as a whole. While the depth of an interaction is crucial for assessing friendship potential, multiple deep interactions may be needed before updating

[19] It is also possible that the higher average number of outgroup participants interacted with at the youth camps (and also for longer hours) in Ghosh et al. (2023) contributed to the effects on friendships observed there, unlike in this setting. Put another way, unlike the work setting, the campers had both broad and deep outgroup contact.

one's perception of the entire group. The results in this section dovetail with this narrative.[20]

Table 7: Effects of contact do not lead to generalized prejudice reduction.

	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***
	(0.08)	(0.12)	(0.08)	(0.06)
Outgroup, deep	-0.08	-0.01	0.20**	0.24***
	(0.09)	(0.10)	(0.09)	(0.07)
Outgroup, broad	-0.04	-0.08	0.14	0.26***
	(0.09)	(0.10)	(0.10)	(0.08)
N	918	918	918	901
Control, raw outcome mean	0.20	276.36	23.25	0.28
Outgroup, deep vs. Outgroup, broad p-value	0.62	0.47	0.49	0.85

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.

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."[21] This study addresses a critical issue: how to bridge deep divides in polarized societies so that 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.

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 schools, 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

[20] 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).

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

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 optimized 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—including those we classify as less visible attributes. As an example, if the most prevalent stereotype is that outgroup members cheat, a production task could be designed in which workers self-report their productivity, without these reports being checked.

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 there is insufficient sample for a rational decision-maker to generalize. Seen in this light, our study may help explain why generalization is seldom seen in the intergroup contact literature. It highlights the need for new, innovative strategies to combat prejudice.

The final takeaway is that majority and minority groups have distinct responses to contact. In our study, contact leads to no updating among the minority group—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 correcting misperceptions should not be a priority area 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.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay, "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking," *Quarterly Journal of Economics*, 2021, *136* (4), 2147–2194.

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.

Banerjee, Abhijit, 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.

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 Pope, "Inaccurate Statistical Discrimination: An Identification Problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The 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.

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.

_ **, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth**, "Justifying dissent," *Quarterly Journal of Economics*, 2023, *138* (3), 1403–1451.

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

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.

Ghosh, Arkadev, "Religious Divisions and Production Technology: Experimental Evidence from India," *Duke University, mimeo*, 2022.

_ **, Prerna Kundu, Matt Lowe, and Gareth Nellis**, "Creating Cohesive Communities: A Youth Camp Experiment in India," *University of British Columbia, mimeo*, 2023.

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

Greene, Kenneth, Erin Rossiter, Enrique Seira, and Alberto Simpser, "Interacting as Equals: How Contact Can Promote Tolerance among Opposing Partisans," *University of Texas at Austin, mimeo*, 2023.

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.

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

Judd, Charles and Bernadette Park, "Definition and Assessment of Accuracy in Social Stereotypes.," *Psychological Review*, 1993, *100* (1), 109.

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?," *University of British Columbia, mimeo*, 2024.

Lyall, Jason, *Divided Armies: Inequality and Battlefield Performance in Modern War*, Princeton: Princeton University Press, 2020.

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

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

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.

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.

Rajkumar, Karthik, Guillaume Saint-Jacques, Iavor Bojinov, Erik Brynjolfsson, and Sinan Aral, "A Causal Test of the Strength of Weak Ties," *Science*, 2022, *377* (6612), 1304–1310.

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.

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.

Thachil, Tariq, "Do Rural Migrants Divide Ethnically in the City? Evidence from an Ethnographic Experiment in India," *American Journal of Political Science*, 2017, *61* (4), 908–926.

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

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 California, San Diego*

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 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 column (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 everyday). 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)	β (3)	p-value (4)	β (5)	p-value (6)	β (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 present 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 column (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 everyday). 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)	(0.02)
Outgroup, deep	0.03	0.01
	(0.03)	(0.01)
Outgroup, broad	0.04	0.01
	(0.03)	(0.01)
N	918	918
Control, mean	0.08	0.01
Outgroup, deep vs. Outgroup, broad *p*-value	0.65	0.85

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 ChatGPT4. 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 $\times$ wave $\times$ day $\times$ shift fixed effects. Robust standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table S7: 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 figure 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 S8: *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 four components: production, identity, bag, and children. The outcome for columns 3, 6, and 9 is the mean of the remaining five components: sanitizer, movie, risk, dictator, and coin. In panel A, all components (prior to standarization) 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 S9: 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).

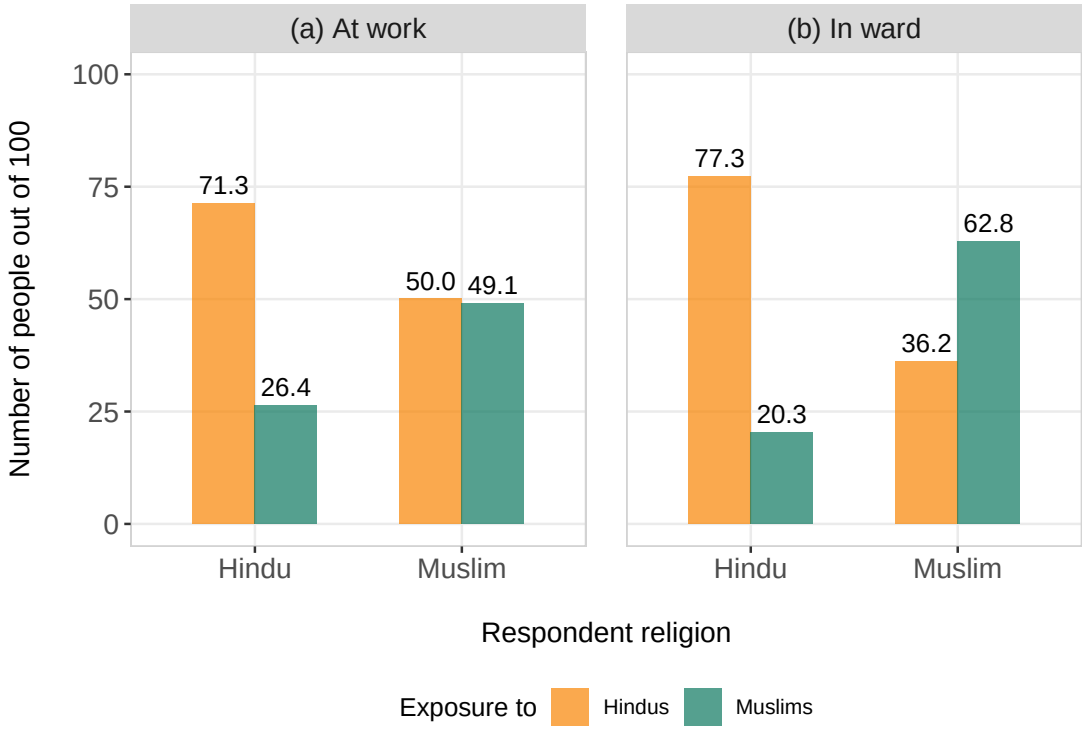
Table S10: No effect of contact on political outcomes.

	Outcome (z-scores):		
	National identity (1)	Democracy rank (2)	Modi thermometer (3)
Ingroup, deep	0.01	0.13	0.06
	(0.08)	(0.12)	(0.05)
Outgroup, deep	0.03	0.15*	0.07
	(0.06)	(0.08)	(0.05)
Outgroup, broad	-0.02	0.16	0.08
	(0.06)	(0.10)	(0.05)
N	918	918	918
Control, raw outcome mean	3.70	1.71	63.24
Outgroup, deep vs. Outgroup, broad p-value	0.49	0.85	0.88

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.

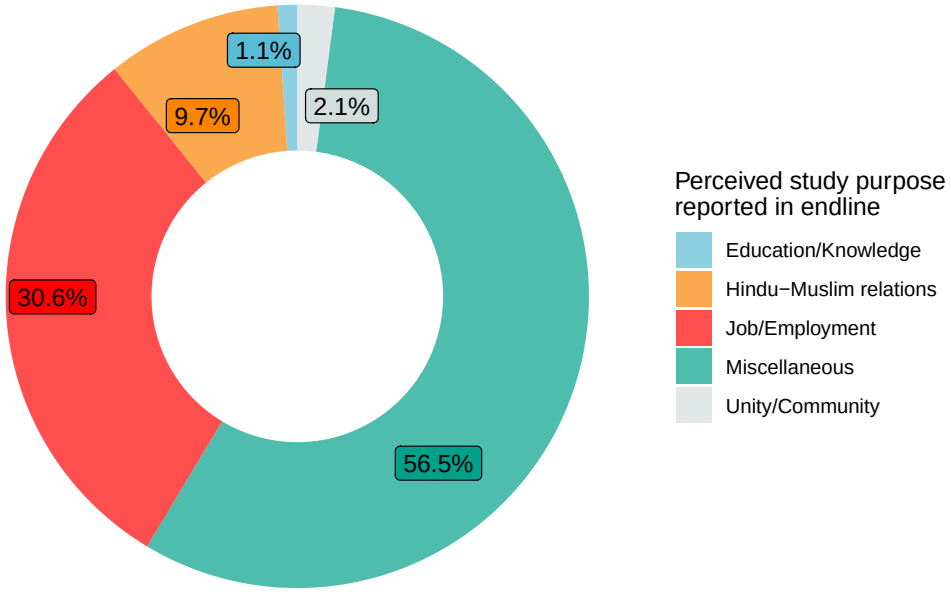
B Supplementary Figures

Figure S1: Exposure to other religious groups at work and in home ward, by respondent religion.



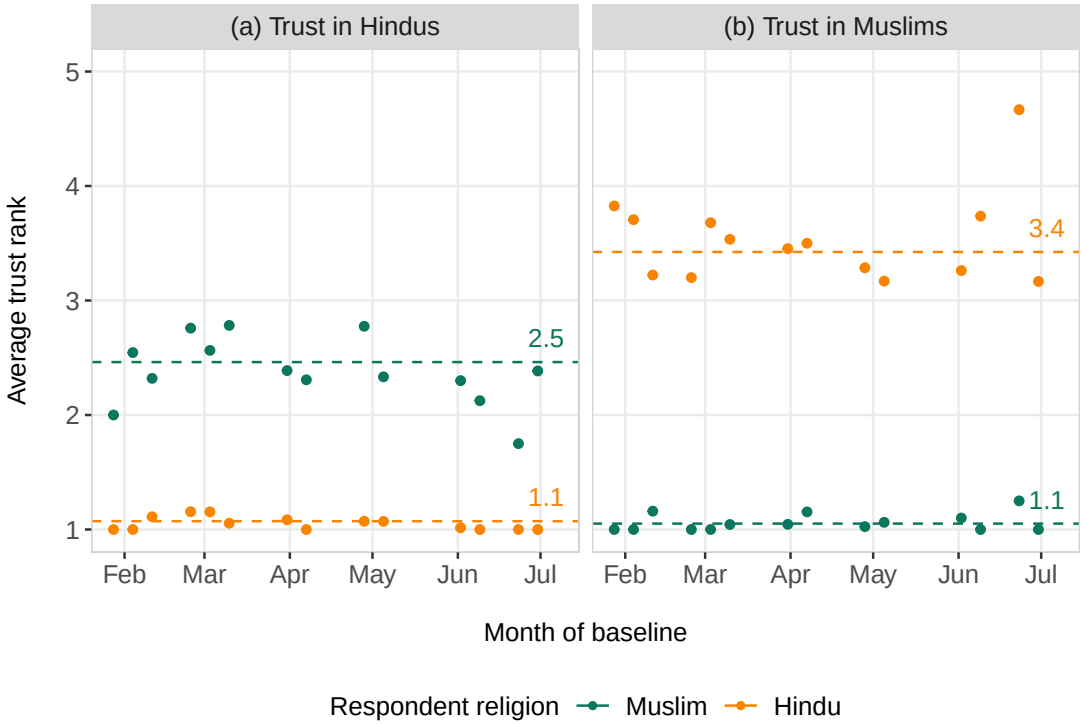
Notes: This figure summarizes how many people from different groups respondents estimate there are in the places where they live and work. The survey questions are: (a) "In your ward, out of every 100 people, roughly how many people belong to each of the following groups?"; (b) "Think about your current 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?" 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 S2: The purpose of the study, according to endline respondents.



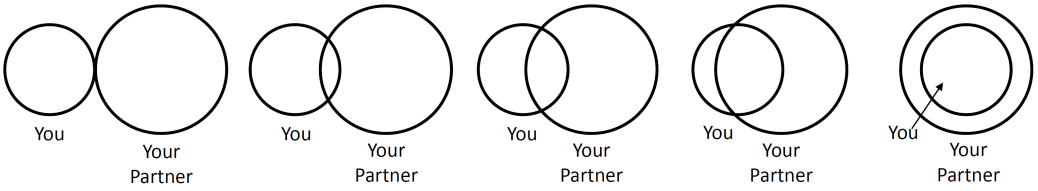
Notes: This figure summarizes endline subjects' reports of the purpose of the study. Subjects were asked at endline, "What do you think was the purpose of this research study? What was it trying to understand?" Subjects qualitative responses were recorded by enumerators and categorized using ChatGPT4.

Figure S3: Baseline trust rankings, by respondent religion.



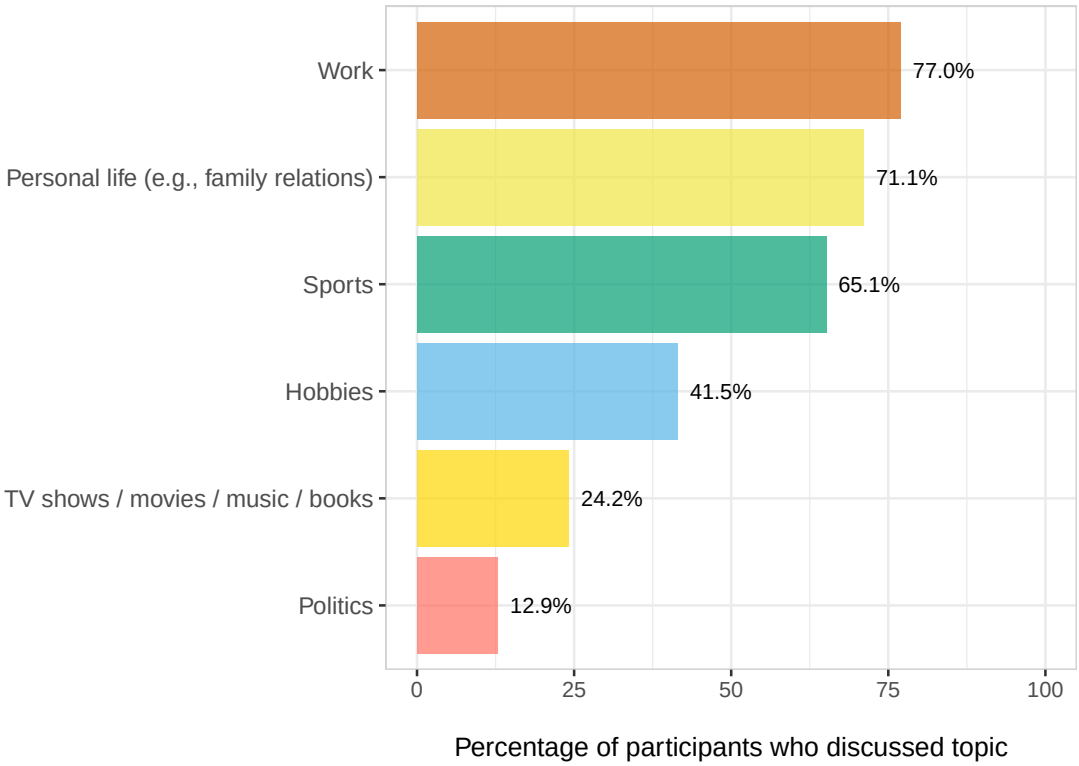
Notes: The dots in this figure show the average trust rankings, over two-week time intervals, for both Hindu and Muslim respondents, across the duration of the study. The data are from the Baseline 1 survey question: "Please rank the following groups in terms of how much you trust people from these groups. 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 S4: Inclusion of partner in self measure, from daily mood survey.



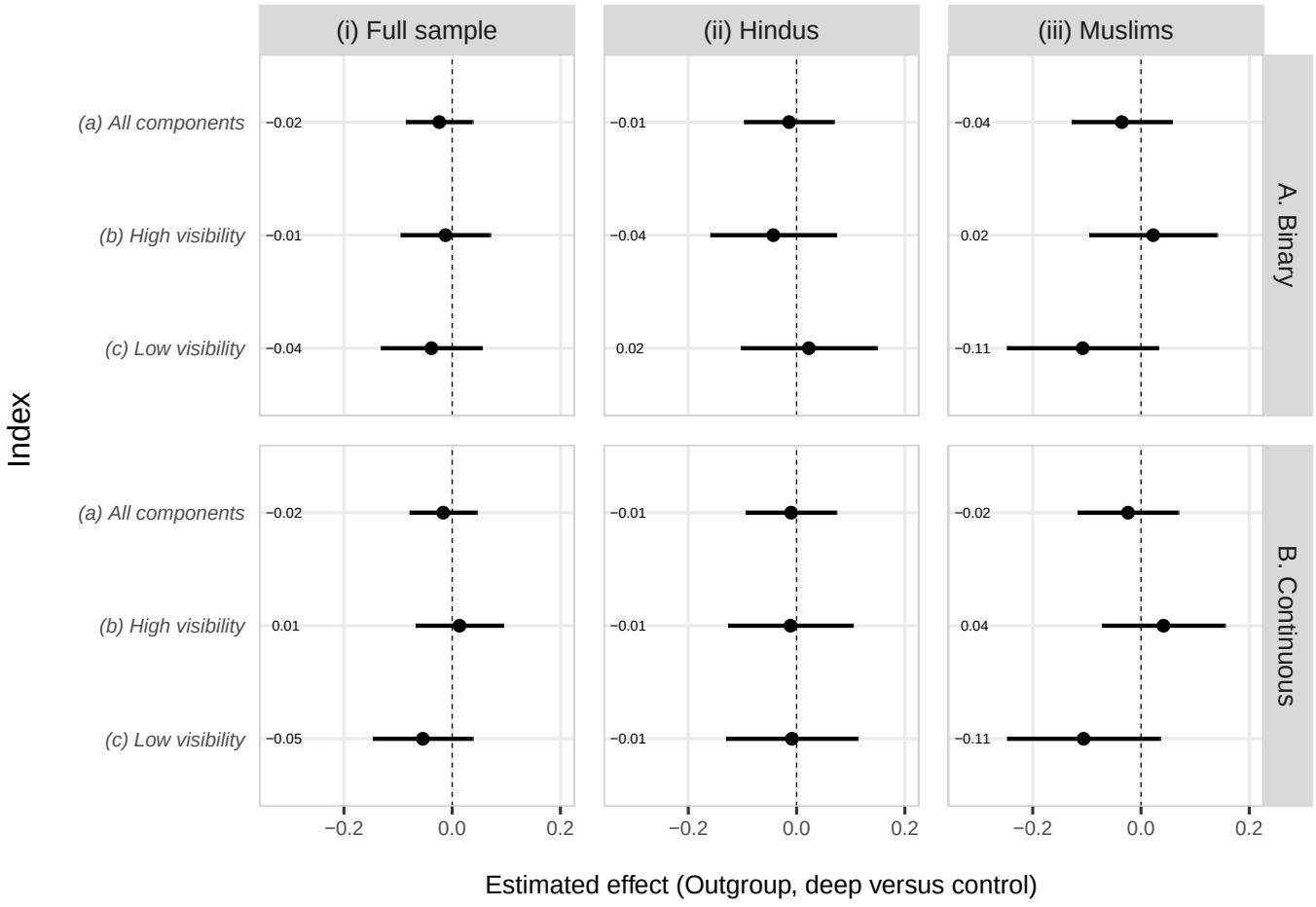
Notes: This figure shows the response options for the following question in the daily mood survey: "Which picture best describes your relationship with your partner today?"

Figure S5: Conversation topics reported on Day 6 of the intervention.



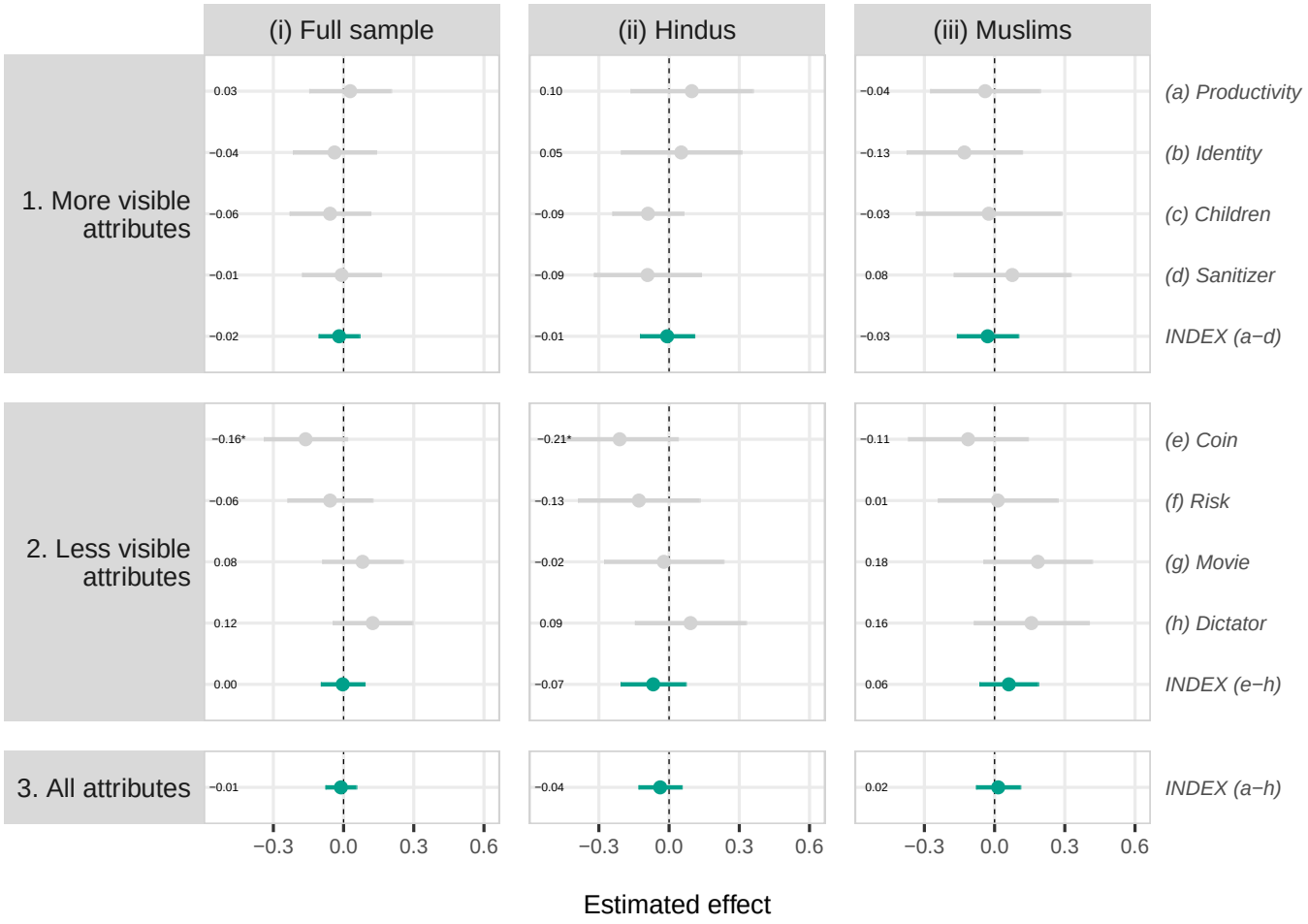
Notes:. This figure shows the percentages of subjects reporting various conversation topics with their partners on Day 6 of the intervention. They are based on the survey question: "Today, what topics did you discuss with your partner? (Select all that apply)." We do not summarize the responses to the "Other (please specify)" option.

Figure S6: *Outgroup, deep* does not increase prediction accuracy about the preferences and behaviors of outgroup strangers relative to the control group.



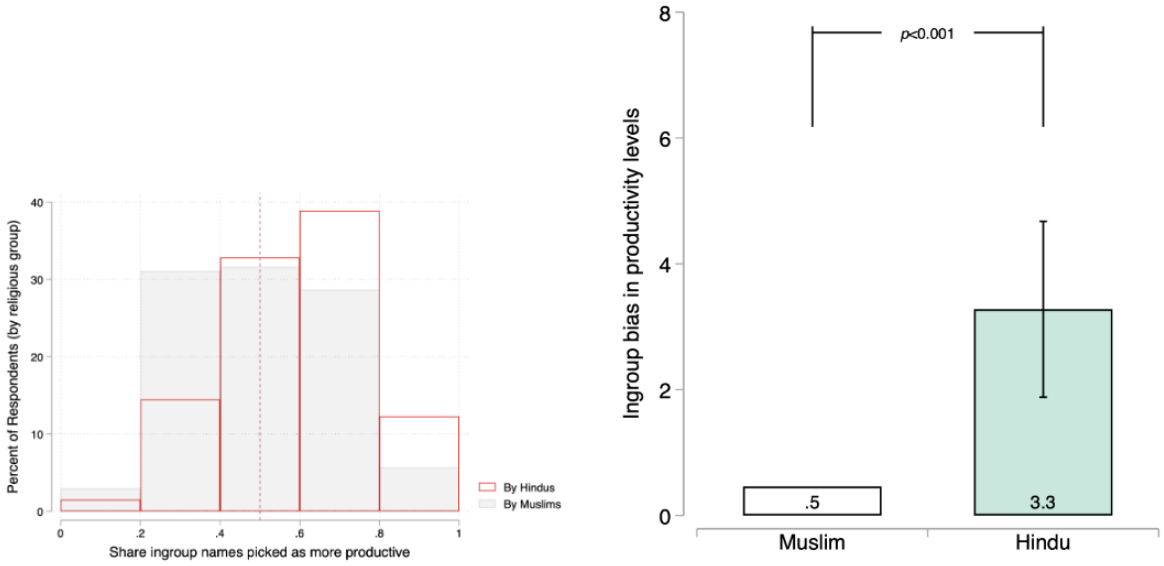
Notes:. This figure parallels Figure 5, but shows the effect of *Outgroup, deep* on the accuracy of guesses about outgroups relative to the pure control group. Each dot and bar summarizes the result of one regression. The sample includes subjects assigned to *Outgroup, deep* or control. The unit of analysis 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 standarization) 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 S7: *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 *Bag* component cannot be included in this figure as this question was not asked about ingroups at endline. 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 *Outgroup, deep* condition. The outcome for panels in row (3) is the mean of the eight components measuring prediction accuracy, as described in the main text. The outcome for panels in row (1) is the mean of the four components: production, 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 S8: Beliefs about the productivity of ingroup and outgroup members.



(a) Extensive Margin

(b) Combining extensive and intensive margins

Note: 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) reports 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) reports the result of sum the intensive margin predictions for cases when the ingroup member is predicted to be more productive minus the sum of predictions for cases when the outgroup member is predicted to be more productive. 95% confidence intervals based on robust standard errors are reported.

Figure S9: 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.

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 exhaustive analysis of these outcomes.

Our only deviation from the pre-registration is that 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).

C.2 Expression of $\widehat{x_{n,D}}^{gen}$ used in Equation 4

$$
\begin{aligned}
\mathscr{Y} &= E\left(\widehat{x}(w^{st},(w_{id})_{i,d}) - x^* - \eta_m\right)^2 \\
&= E\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^* - \eta_m\right)^2 \\
&= E\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^*\right)^2 - 2E\left(\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^*\right)\eta_m\right) + (\eta_m^2) \\
&= E\left(w^{st}(x_0^{st} - x^*) + \sum_{i\le n}\sum_{d\le D} w_{id}(x_{id} - x^*)\right)^2 + 0 + \sigma_{het}^2 \\
&= E\left(w^{st}(\varepsilon_0^{st}) + \sum_{i\le n}\sum_{d\le D} w_{id}(\eta_i + \varepsilon_{id})\right)^2 \\
&= (w^{st})^2 E(\varepsilon_0^{st})^2 + \sum_{i\le n}\sum_{d\le D}(w_{id})^2 E(\eta_i^2 + \varepsilon_{id}^2) + \sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})E(\eta_i^2) + \sigma_{het}^2 \\
&= (w^{st})^2\sigma_{st}^2 + \sum_{i\le n}\sum_{d\le D}(w_{id})^2(\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})\sigma_{het}^2 + \sigma_{het}^2 \qquad (5) \\
&= (1 - \sum_{i\le n}\sum_{d\le D} w_{id})^2\sigma_{st}^2 + \sum_{i\le n}\sum_{d\le D}(w_{id})^2(\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})\sigma_{het}^2 + \sigma_{het}^2 \qquad (6)
\end{aligned}
$$

[22]

To derive the expression in (5), 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 (6) with respect to any $w_{i^*d^*}$ is:

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

The second derivative is positive, implying we are at a minima. Given we obtain the same first derivative for all i,d, we use symmetry to impose $w_{id} = w$ and then obtain

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

[22] The sum $\sum_{d>d'}$ implies summing over all ordered pairs d,d' where $d \neq d'$. The "ordered" feature helps in accounting.

C.3 Expression of $\widehat{x_{n,D}}^I$

$$
\begin{aligned}
\mathscr{V} &= E\left(\widehat{x}(w^{st},(w_{id})_{i,d}) - x^* - \eta_I\right)^2 \\
&= E\left(w^{st}x_0^{st} + \sum_{i\le n}\sum_{d\le D} w_{id}x_{id} - x^* - \eta_I\right)^2 \\
&= E\left(w^{st}(x_0^{st} - x^* - \eta_I) + \sum_{i\le n}\sum_{d\le D} w_{id}(x_{id} - x^* - \eta_I)\right)^2 \\
&= E\left(w^{st}(\varepsilon_0^{st} - \eta_I) + \sum_{i\neq I}\sum_{d\le D} w_{id}(\eta_i + \varepsilon_{id} - \eta_I) + \sum_{d\le D} w_{Id}\varepsilon_{Id}\right)^2 \\
&= (w^{st})^2 E\left((\varepsilon_0^{st})^2 + \eta_I^2\right) + \sum_{i\neq I}\sum_{d\le D}(2w^{st}w_{id})E(\eta_I^2) + \sum_{i\neq I}\sum_{d\le D}(w_{id})^2 E(\eta_i^2 + \varepsilon_{id}^2 + \eta_I^2) + \sum_{i\neq I}\sum_{d\neq d'}(2w_{id}w_{id'})E(\eta_i^2 + \eta_I^2) \\
&\quad + \sum_{i\neq i';i,i'\neq I}\sum_{d,d'\le D}(2w_{id}w_{i'd'})\eta_n^2 + \sum_{d\le D} w_{Id}^2\varepsilon_{Id}^2 \\
&= (w^{st})^2(\sigma_{st}^2 + \sigma_{het}^2) + 2w^{st}\sum_{i\neq I}\sum_{d\le D} w_{id}\sigma_{het}^2 + \sum_{i\neq I}\sum_{d\le D}(w_{id})^2(2\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{i\neq I}\sum_{d\neq d'}(2w_{id}w_{id'})2\sigma_{het}^2 \\
&\quad + \sum_{i\neq i';i,i'\neq I}\sum_{d,d'\le D}(2w_{id}w_{i'd'})\sigma_{het}^2 + \sum_{d\le D} w_{Id}^2\sigma_{inf}^2 \qquad (7)
\end{aligned}
$$

Replacing $w^{st} = (1 - \sum_{i\le n}\sum_{d\le D} w_{id})$ and taking the first derivative with respect to any $w_{i^*d^*}, i^* \neq I$ gives:

$$
\begin{aligned}
-2(1 - \sum_{i\le n}\sum_{d\le D} w_{id})(\sigma_{st}^2 + \sigma_{het}^2) - 2\sum_{i\neq I}\sum_{d\le D} w_{id}\sigma_{het}^2 + 2(1 - \sum_{i\le n}\sum_{d\le D} w_{id})\sigma_{het}^2 + 2w_{i^*d^*}(2\sigma_{het}^2 + \sigma_{inf}^2) + \sum_{d\neq d^*}(2w_{i^*d})2\sigma_{het}^2 \\
+ \sum_{i\neq i^*, i\neq I}\sum_{d\le D}(2w_{id})\sigma_{het}^2 = 0
\end{aligned}
$$

Similarly, taking the first derivative with respect to any w_{nd^*} gives:

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

The second derivatives are positive, implying we are at a minima. Given we obtain the symmetric derivatives for all $w_{id}, i \neq n$ and for all w_{Id}, we use symmetry to impose $w_{id} = w_1$ for $i \neq n$ and $w_{Id} = w_2$, and get the two equations:

$$
\begin{aligned}
-2(1 - (n-1)Dw_1 - Dw_2)\sigma_{st}^2 - 2(n-1)Dw_1\sigma_{het}^2 + 2(1 - (n-1)Dw_1 - Dw_2)\sigma_{het}^2 + 2w_1(2\sigma_{het}^2 + \sigma_{inf}^2) + 4(D-1)w_1\sigma_{het}^2 \\
+(n-2)D(2w_1)\sigma_{het}^2 = 0 \\
-2(1 - (n-1)Dw_1 - Dw_2)\sigma_{st}^2 - 2(n-1)Dw_1\sigma_{het}^2 + 2w_2\sigma_{inf}^2 = 0
\end{aligned}
$$

This solves as

$$
w_1 = \frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2 + D\sigma_{het}^2)(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)}, w_2 = \frac{\sigma_{st}^2\sigma_{inf}^2 + nD\sigma_{st}^2\sigma_{het}^2 + \sigma_{inf}^2\sigma_{het}^2 + D\sigma_{het}^4}{(\sigma_{inf}^2 + D\sigma_{het}^2)(nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2)}, w^{st} = \frac{\sigma_{inf}^2}{nD\sigma_{st}^2 + \sigma_{inf}^2 + D\sigma_{het}^2}
$$

C.4 Omitted proof for Result 1

Proof. Part (i): In $\mathscr{V}$ derived in Equation 5, we change $E(x_0^{st}-x^*)^2$ from $E(\varepsilon_0^{st})^2=\sigma_{st}^2$ to simply $(x_0^{st}-x^*)^2$, to get:

$$\begin{aligned}
\mathscr{V}^{gen} &= (w^{st})^2(x_0^{st}-x^*)^2+\sum_{i\le n}\sum_{d\le D}(w_{id})^2E(\eta_i^2+\varepsilon_{id}^2)+\sum_{i\le n}\sum_{d>d'}(2w_{id}w_{id'})E(\eta_i^2)+\sigma_{het}^2\\
&= \left(\frac{\sigma_{inf}^2+D\sigma_{het}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2(x_0^{st}-x^*)^2+\underbrace{nD}_{N}\left(\frac{\sigma_{st}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2(\sigma_{het}^2+\sigma_{inf}^2)\\
&\quad+\underbrace{nD}_{N}(D-1)\left(\frac{\sigma_{st}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2\sigma_{het}^2+\sigma_{het}^2\\
&= \frac{(\sigma_{inf}^2+D\sigma_{het}^2)^2b^2+nD\sigma_{st}^4\left((\sigma_{het}^2+\sigma_{inf}^2)+(D-1)\sigma_{het}^2\right)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}+\sigma_{het}^2\\
&= (\sigma_{inf}^2+D\sigma_{het}^2)\frac{(\sigma_{inf}^2+D\sigma_{het}^2)b^2+nD\sigma_{st}^4}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}+\sigma_{het}^2\\
\frac{\partial\mathscr{V}^{gen}}{\partial n} &= -(\sigma_{inf}^2+D\sigma_{het}^2)\frac{D\sigma_{st}^2\left(2b^2\sigma_{inf}^2+2Db^2\sigma_{het}^2+nD\sigma_{st}^4-\sigma_{st}^2\sigma_{inf}^2-D\sigma_{st}^2\sigma_{het}^2\right)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^3}
\end{aligned}$$

$\frac{\partial\mathscr{V}^{gen}}{\partial n}<0$ is guaranteed by $2b^2\ge\sigma_{st}^2$ and hence also by $3b^2\ge2\sigma_{st}^2$.

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

$\frac{d^2\mathscr{V}^{gen}}{dn^2}>0$ is guaranteed by $3b^2\ge2\sigma_{st}^2$.

Next, we calculate the comparative statics w.r.t D. For the first derivative with respect to D:

$$\begin{aligned}
\frac{\partial\mathscr{V}^{gen}}{\partial D} &= \frac{b^2\sigma_{het}^2+n\sigma_{st}^4}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^2}-\frac{2\left(b^2(\sigma_{inf}^2+D\sigma_{het}^2)+Dn\sigma_{st}^4\right)(n\sigma_{st}^2+\sigma_{het}^2)}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^3}\\
&= \frac{-2b^2n\sigma_{st}^2\sigma_{inf}^2-Db^2n\sigma_{st}^2\sigma_{het}^2-b^2\sigma_{inf}^2\sigma_{het}^2-Db^2\sigma_{het}^4-Dn^2\sigma_{st}^6+n\sigma_{st}^4\sigma_{inf}^2-Dn\sigma_{st}^4\sigma_{het}^2}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^3}
\end{aligned}$$

For the second derivative with respect to D:

$$\frac{\partial^2\mathscr{V}^{gen}}{\partial D^2}=\frac{\sigma_{st}^2\begin{pmatrix}4Db^2n\sigma_{st}^2\sigma_{inf}^2+2D^2b^2n\sigma_{st}^2\sigma_{het}^2-2b^2\sigma_{inf}^4+2D^2b^2\sigma_{het}^4+D^2n^2\sigma_{st}^6\\-4Dn\sigma_{st}^4\sigma_{inf}^2+\sigma_{st}^2\sigma_{inf}^4-D^2\sigma_{st}^2\sigma_{het}^4\end{pmatrix}}{(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)^4}$$

As before, $2b^2\ge\sigma_{st}^2$ guarantees that $\frac{\partial\mathscr{V}^{gen}}{\partial D}<0$. $\frac{\partial^2\mathscr{V}^{gen}}{\partial D^2}>0$ is still not guaranteed without additional assumptions.

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

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

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

In $\mathscr{V}^I$ derived in Equation 7, we change $E(x_0^{st} - x^*)^2$ from $E(\varepsilon_0^{st})^2 = \sigma_{st}^2$ to simply $(x_0^{st} - x^*)^2$, to get:

$$\begin{aligned}
\mathscr{V}^I &= (w^{st})^2(b^2+\sigma_{het}^2) + 2w^{st}\sum_{i\neq I}\sum_{d\leq D} w_{id}\sigma_{het}^2 + \sum_{i\neq I}\sum_{d\leq D}(w_{id})^2(2\sigma_{het}^2+\sigma_{inf}^2) + \sum_{i\neq I}\sum_{d\neq d'}(2w_{id}w_{id'})2\sigma_{het}^2 \\
&\quad + \sum_{i\neq i';i,i'\neq I}\sum_{d,d'\leq D}(2w_{id}w_{i'd'})\sigma_{het}^2 + \sum_{d\leq D} w_{Id}^2\sigma_{inf}^2 \\
&= \left(\frac{\sigma_{inf}^2}{nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2}\right)^2(b^2+\sigma_{het}^2) + (n-1)D\left(\frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)}\right)^2(2\sigma_{het}^2+\sigma_{inf}^2) \\
&\quad + (n-1)D(D-1)\left(\frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)}\right)^2 2\sigma_{het}^2 \\
&\quad + n(n-1)D^2\left(\frac{\sigma_{st}^2\sigma_{inf}^2}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)}\right)^2\sigma_{het}^2 + D\left(\frac{\sigma_{st}^2\sigma_{inf}^2+nD\sigma_{st}^2\sigma_{het}^2+\sigma_{inf}^2\sigma_{het}^2+D\sigma_{het}^4}{(\sigma_{inf}^2+D\sigma_{het}^2)(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2)}\right)^2\sigma_{inf}^2
\end{aligned}$$

Now that we have the expressions of variance under both conditions, here is a property we implicitly used while defining $\mathscr{G}(n,D) = \frac{\mathscr{A}^{gen}}{\mathscr{A}^I}$:

$$\mathscr{V}^{gen} - \mathscr{V}^I = \frac{D\sigma_{het}^2\begin{bmatrix} 2b^2\sigma_{inf}^6 + 5Db^2\sigma_{inf}^4\sigma_{het}^2 + 4D^2b^2\sigma_{inf}^2\sigma_{het}^4 + D^3b^2\sigma_{het}^6 + D^2n^2\sigma_{st}^4\sigma_{inf}^2\sigma_{het}^2 \\ +D^3n^2\sigma_{st}^4\sigma_{het}^4 + 3D^2n\sigma_{st}^4\sigma_{inf}^2\sigma_{het}^2 + D^3n\sigma_{st}^4\sigma_{het}^4 + 2n\sigma_{st}^2\sigma_{inf}^6 + 4Dn\sigma_{st}^2\sigma_{inf}^4\sigma_{het}^2 \\ +4D^2n\sigma_{st}^2\sigma_{inf}^2\sigma_{het}^4 + 2D^3n\sigma_{st}^2\sigma_{het}^6 + 2D\sigma_{st}^4\sigma_{inf}^4 - 2\sigma_{st}^2\sigma_{inf}^6 - 2D\sigma_{st}^2\sigma_{inf}^4\sigma_{het}^2 \\ +\sigma_{inf}^6\sigma_{het}^2 + 3D\sigma_{inf}^4\sigma_{het}^4 + 3D^2\sigma_{inf}^2\sigma_{het}^6 + D^3\sigma_{het}^8 \end{bmatrix}}{\left(\sigma_{inf}^2+D\sigma_{het}^2\right)^2\left(nD\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2\right)^2} > 0$$

$$\frac{d\mathscr{V}}{dD}|_{N=nD} = -\frac{\sigma_{inf}^2\sigma_{het}^2\begin{pmatrix} N^3\sigma_{st}^6\sigma_{inf}^2\sigma_{het}^2 + DN^3\sigma_{st}^6\sigma_{het}^4 - N^2\sigma_{st}^6\sigma_{inf}^4 + 3DN^2\sigma_{st}^6\sigma_{inf}^2\sigma_{het}^2 + N^2\sigma_{st}^4\sigma_{inf}^4\sigma_{het}^2 \\ +4DN^2\sigma_{st}^4\sigma_{inf}^2\sigma_{het}^4 + 3D^2N^2\sigma_{st}^4\sigma_{het}^6 + 4DN\sigma_{st}^6\sigma_{inf}^4 - N\sigma_{st}^4\sigma_{inf}^6 + 8DN\sigma_{st}^4\sigma_{inf}^4\sigma_{het}^2 \\ +9D^2N\sigma_{st}^4\sigma_{inf}^2\sigma_{het}^4 - 3N\sigma_{st}^2\sigma_{inf}^6\sigma_{het}^2 - 3DN\sigma_{st}^2\sigma_{inf}^4\sigma_{het}^4 + 3D^2N\sigma_{st}^2\sigma_{inf}^2\sigma_{het}^6 \\ +3D^3N\sigma_{st}^2\sigma_{het}^8 + 2b^2\sigma_{inf}^8 + 6Db^2\sigma_{inf}^6\sigma_{het}^2 + 6D^2b^2\sigma_{inf}^4\sigma_{het}^4 + 2D^3b^2\sigma_{inf}^2\sigma_{het}^6 \\ +4D\sigma_{st}^4\sigma_{inf}^6 - 4D^3\sigma_{st}^4\sigma_{inf}^2\sigma_{het}^4 - 2\sigma_{st}^2\sigma_{inf}^8 + 6D^2\sigma_{st}^2\sigma_{inf}^4\sigma_{het}^4 \\ +4D^3\sigma_{st}^2\sigma_{inf}^2\sigma_{het}^6 + \sigma_{inf}^8\sigma_{het}^2 + 4D\sigma_{inf}^6\sigma_{het}^4 + 6D^2\sigma_{inf}^4\sigma_{het}^6 \\ +4D^3\sigma_{inf}^2\sigma_{het}^8 + D^4\sigma_{het}^{10} \end{pmatrix}}{\left(\sigma_{inf}^2+D\sigma_{het}^2\right)^3\left(N\sigma_{st}^2+\sigma_{inf}^2+D\sigma_{het}^2\right)^3}$$

By comparing the similarly colored terms, roughly grouped together based on the power of σ_{st}, one can easily see that $\sigma_{het}^2 \geq \sigma_{inf}^2$ guarantees $\frac{d\mathscr{V}}{dD}|_{N=nD} < 0$ under $N \geq D \geq 1$.

C.5 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.6 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?