Silence to Solidarity: How Communication About a Minority Affects Discrimination

Duncan Webb*

December 29, 2025

Download latest version

Abstract

People predominantly talk within their identity groups. Conversations between majority-group members may therefore be crucial in shaping discrimination against minorities. In an experiment in India (N=3,397), non-transgender participants strongly discriminate against transgender workers when hiring for a grocery delivery, but a discussion with two non-transgender neighbors eliminates this discrimination in private post-discussion choices. The discussion is 1.7x more effective at reducing discrimination than information about transgender people's legal rights, and effects partially persist after 1 month. The evidence is consistent with norm-based persuasion, in which pro-transgender participants are more vocal and create a perceived anti-discriminatory norm that reduces subsequent discrimination.

JEL Codes: *J15, D83, J71, C93, K38, Z13*

Keywords: *discrimination, communication, social interactions, transgender, legal rights, persuasion*

*Nova School of Business and Economics, dmbwebb@gmail.com. I'd like to thank my supervisors Karen Macours and Suanna Oh for all their fantastic advice and support, and to Abhijit Banerjee, Esther Duflo, Supreet Kaur, and Frank Schilbach for their invaluable guidance throughout the project. I'd also like to thank 5 anonymous reviewers, along with Peter Andre, Sam Asher, David Atkin, Luc Behaghel, David Bernard, Francis Bloch, Leonardo Bursztyn, Janet Currie, Denis Cogneau, Pascaline Dupas, Oliver Vanden Eynde, Evan Friedman, Amory Gethin, Julien Grenet, Deivy Houeix, Nicolas Jacquemet, Seema Jayachandran, Eliana La Ferrara, Sylvie Lambert, Elisa Macchi, David Margolis, Edward Miguel, Eduardo Montero, Oda Nedregård, Kate Orkin, Garima Sharma, Morten Nyborg Støstad, Nick Otis, Tom Raster, Kailash Rajah, Gautam Rao, Matthew Ridley, James Robinson, Chris Roth, Raul Sanchez de la Sierra, Advik Shreekumar, Tavneet Suri, Simon Quinn, Jean-Marc Tallon, Lore Vandewalle, Liam Wren-Lewis, and many seminar participants at PSE, MIT, Berkeley, NEUDC, AFEDEV, the CEPR forum, and AFE for helpful comments and feedback. Akhilan Rengaswamy, Arun Balaji, Bhala Dhapandani, and Manvi Govil were invaluable in coordinating the fieldwork. This work has been generously supported by UK International Development, awarded through the J-PAL Crime and Violence Initiative, EUR-PgSE, CEPREMAP, the Development Research Group at PSE, the Weiss Fund, the Institute for Humane Studies (grant # IHS017466), and Fundação para a Ciência e a Tecnologia (UID/00124/2025, UID/PRR/124/2025, Nova School of Business and Economics), and LISBOA2030 (DataLab2030 - LISBOA2030-FEDER-01314200). All errors are my own. This study was pre-registered in the AEA RCT Registry under the unique identifying number AEARCTR-0010953. The study was approved by the Institutional Review Board at both Paris School of Economics and the Institute for Financial Management, Chennai.

1 Introduction

Discriminatory behavior harms equity and efficiency in a wide range of economic domains, including in firms (**???**), the labor market (**????**), housing (**?**), and informal social interactions (**?**). Standard theories of discrimination frame it as an individual's decision, based on beliefs or deep-seated preferences that are difficult to change (**????**). But at the societal level, we sometimes observe rapid changes in discriminatory preferences (**???**). For example, multiple countries have seen rapid increases in the proportion of people who accept interethnic marriage, homosexuality, and equal rights for women over the course of a single generation.[1] Some of these changes may be driven by social contact *between* majority and minority groups (**??**). But in a world of homophily, where people mostly interact with others within their identity group (**?**), communication *within* the majority group may also be a powerful determinant of change.

In this paper, I therefore test how communication between majority-group members about a minority ("*horizontal communication*") affects discrimination. Standard models typically predict that communication merely averages out individual beliefs, leading beliefs to become more similar within a group, but not changing the mean belief or resulting average behavior.[2] Yet horizontal communication could lead to large shifts in post-communication discrimination if there are asymmetries between pro- and anti-minority communication. For example, pro-minority individuals might be more persuasive, using moral arguments in their favor; anti-minority people might have stronger preferences against minorities, and so choose to speak up more in discussions; or anti-minority individuals may be unwilling to express discriminatory views if they are concerned about how they are perceived by others. Such asymmetries could drive changes in post-communication discrimination if communication affects people's attitudes, or beliefs, or the norm of behavior towards a minority.

I run a field experiment in urban Chennai, India (N=3397) that tests these ideas in the context of discrimination against the most visible LGBTQ+ group in India: a community of transgender women known locally as *thirunangai*. This setting is an appropriate context in which to study the effect of communication on discrimination. The community is vulnerable to extensive economic discrimination and violence (**?**), and their distinct visual identity and historic role in Indian society make them highly recognizable – allowing me to measure discrimination only by showing photos. At the same time, there appears to be nascent social change towards greater acceptance of transgender people. This may create conditions in which communication reduces discrimination: for example, despite strong *de facto* discrimination, there is also widespread agreement that discrimination is wrong.[3] This raises the possibility that pro-trans individuals

[1] For example: (i) the proportion of people in the US saying that homosexuality is wrong dropped rapidly in a single year (1993) and continued to decline quickly in the 30 following years (**?**); (ii) the proportion of people in the UK indicating discomfort with interethnic marriage dropped from 55% in 1983 to 25% in 2013 (**?**); and (iii) the proportion of people in Uganda saying that women should have equal rights went from 63% in 2002 to 80% in 2012 (**?**).

[2] For example, agents who truthfully share information will eventually converge to the same posterior belief in standard Bayesian models (**??**). And if agents' priors are not systematically incorrect, the martingale property of Bayesian updating implies that their beliefs should not systematically shift in one direction.

[3] Despite observing substantial hiring discrimination in my control group, 93% of that same control group say that discrimination is unacceptable in response to a vignette that showcases explicit discrimination. There appears, therefore, to be a wedge between the descriptive norm (how much people actually discriminate) and the prescriptive norm (to what extent people think it is right or wrong to discriminate).

may be more vocal or more persuasive in such settings.

I first evaluate whether horizontal communication can affect discrimination by randomizing whether participants are involved in a group discussion with two of their neighbors. I measure the effect of this discussion on anti-transgender discrimination in a series of private, individual hiring choices after the discussion has ended. Participants are offered a free grocery delivery, and make a series of binary choices over the worker who will carry out the delivery (along with the items they will receive, which are randomly varied across choices). Participants who do not take part in a discussion are highly discriminatory: they are 19 percentage points (32%) less likely to hire transgender workers than non-transgender workers ($p<0.001$). Their choices imply that they are willing to sacrifice grocery items worth 1.9x the median daily food expenditure to avoid interacting with a transgender worker for 15 minutes.

Horizontal communication leads to large short-run reductions in subsequent discrimination: in private choices after a discussion, there is no discrimination against transgender workers on average. In the discussion condition, participants discuss a series of hiring options as a group of three neighbors and are asked to make collective hiring choices.[4,5] Since some of these options include transgender workers, participants naturally discuss whether to hire transgender workers. Crucially, the only communication about transgender people in this discussion comes from the participants themselves, rather than from the discussion facilitator.

The effects of this discussion on discrimination are stark: in people's private, post-discussion hiring choices, participants are 17 p.p. (42%) more likely to select a transgender worker than the control group ($p<0.001$), implying that anti-transgender discrimination is reduced to 0 on average (p of difference between transgender and non-transgender: 0.30). The effects are also partially persistent, although there is substantial fadeout: when I re-survey participants approximately 1 month later, discussion participants are 4 p.p. more likely to select transgender workers than the control group in a series of hypothetical hiring choices ($p=<0.001$).

I then compare the effects of the discussion with communication about a minority from an agent of authority, namely, the legal system ("*top-down communication*"). I cross-randomize whether participants watch a video informing them about an Indian Supreme Court ruling that affirmed that transgender people have all the same fundamental rights as other citizens, including freedom from discrimination.[6] Such communication could affect discrimination if it signals that discrimination is not socially acceptable (**???**) or if participants fear legal enforcement. Learning about these legal rights also lowers discrimination by 10.3 p.p. ($p<0.001$). But the effects are only 59% as large as the effects of the group discussion (p of difference < 0.05),

[4]Including hiring choices in the discussion reduces the possibility of cheap talk by increasing the stakes of the discussion. It also implies that I measure the bundled effect of (i) observing others' choices and (ii) listening to the arguments made in the discussion. I later show evidence that both matter: observing others' choices without a discussion moderately reduces discrimination, while features like how much *morality* was invoked in the arguments correlate with reduced discrimination even when controlling for choices.

[5]Groups were always same-gender to avoid overly hierarchical relationships between group members. To test for the effects of *within-network* horizontal communication, to make the discussions more naturalistic, and to ensure social image concerns could affect behavior, we recruited neighbors who knew each other 98% of the time.

[6]There are no interaction effects between the legal rights video and the group discussion ($p \in [0.96, 1.00]$). The effect of the discussion is also not driven by interaction effects; it is robust to only using the sample who were not informed about transgender rights.

and do not persist after 1 month. In this context, therefore, horizontal communication about transgender people is substantially more effective than top-down information about the law.

I then seek to understand why discussions reduce discrimination in this context. The suggestive evidence is consistent with *norm-based persuasion* as the primary mechanism, in which pro-transgender participants speak up more during discussions, use especially persuasive moral arguments, and thus generate a perception of a strong anti-discriminatory norm that persuades others to discriminate less in their subsequent private choices.[7]

I first use a mechanism treatment to show the effects are driven by interpersonal persuasion rather than self-persuasion (**?**). Silent "listeners", who listen to but do not participate in discussions, discriminate 13 p.p. less afterward ($p<0.001$) – an effect statistically indistinguishable from that of active participants. This finding supports interpersonal persuasion since listeners do not speak in the discussion and yet still show treatment effects as strong as participants. Consistent with this, the more pro-transgender statements a listener hears from others, the less she subsequently discriminates.

Second, I use correlational evidence to show that persuasion appears to operate through changes in perceived descriptive norms – shifts in what people expect others to do. Norm updates alone suffice to reduce discrimination: in a treatment without discussions, "observers" learn their group members' first-round choices, shift their beliefs about within-group norms, and subsequently discriminate less. Specifically, participants predict that their groups will privately discriminate 4.5 p.p. less if they have observed others' choices, and this small norm update translates one-for-one into a 4.5 p.p. reduction in discrimination in observers' own private second-round choices. This direct relationship between perceived norms and post-discussion discrimination confirms that updating perceived descriptive norms—even without the rich interaction of discussion—can reduce discrimination. I then show that the discussion generates large shifts in these perceived norms (24 p.p, $p<0.001$), while producing minimal changes in attitudes or beliefs about transgender workers. Since these norm shifts strongly mediate the discussions' effects on discrimination, this is consistent with norm-updates being a key driver of the discussion's effects.

Norm-based persuasion can explain how participants influence each other's private behavior, but not why discussions create a large mean shift *away* from discrimination in this context. Standard communication models predict that communication should "average out" behavior – behavior should converge within a group, with no systematic change in mean behavior (see Section **??**). The substantial reduction in discrimination therefore requires sources of *asymmetry*, whereby pro-transgender and anti-transgender forces do not simply cancel out. I use correlational evidence to document three potential sources of such asymmetry.

First, privately pro-trans participants speak up more frequently during discussions. Those who privately select more transgender workers after discussions (a proxy for pro-trans pri-

[7]The implied *persuasion rate* (29%) falls at the upper end of the range in the literature (**?**), but remains consistent with other successful in-person interventions and is plausible given the short-term nature of the effects.

vate preferences)[8] are 32% more likely to speak first and 27% more likely to dominate for discussions involving transgender workers. This asymmetric behavior appears to be motivated by a particular desire to influence others, rather than by social image concerns – pro-trans participants are more likely to self-report intentions to shape others' choices than other participants. Pro-trans statements thus occur 2.5x more frequently than anti-trans statements in the transcripts. This is also driven by the substantial minority of pro-trans participants: despite strong average discrimination, 31% of control participants still have a positive willingness to pay to select a transgender worker.

Second, I show suggestive evidence of asymmetric inference about norms: participants more readily shift toward a perception of pro-trans norms than away from them. Discussions *generate* misperceptions – participants overestimate how pro-trans their group members are after a discussion, reversing the underestimation seen in the control group. This suggests that participants fail to account for pro-trans participants' greater propensity to speak up, thus overestimating the prevalence of pro-trans behavior. Descriptive evidence also suggests that participants exhibit *silence neglect*, updating beliefs when hearing explicit pro-trans statements but neglecting the information contained in silence. Participants who hear *no* pro-transgender statements during discussions show no change in perceived norms or behavior, suggesting that silence carries little informational weight, even though the control group's results show that silent participants are largely anti-trans. These patterns are consistent with a "spiral of silence" (**??**), in which selective communication leads people to perceive a view as more prevalent than it truly is. They also link to work on misperceived social norms (**?**), but with a distinctive feature—communication itself generates the misperception.

Finally, I analyze the discussion transcripts to unpack the black box of the discussion and show suggestive evidence of *rhetorical asymmetry* (**?**), in which some arguments are more persuasive than others. In particular, the evidence is consistent with pro-trans arguments being more persuasive than anti-trans arguments because they more often invoke moral language and principles. Participants frequently invoke moral concepts when discussing transgender workers, using language about "opportunities," "rights," and "equality." These moral arguments are correlated with both larger changes in perceived norms and greater reductions in discrimination, even after controlling for within-discussion choices.[9] These findings align with research on moral reframing, which shows that arguments framed in moral terms are especially persuasive when they resonate with listeners' values (**???**).

Why does horizontal communication reduce discrimination here, even though it can clearly promote it in other settings (e.g., **??**)? While I cannot definitively identify which contextual features are essential in the absence of cross-context comparisons, the mechanism results suggest one crucial feature: the stark disconnect between the descriptive norm (widespread

[8] I did not include baseline measures of discrimination in order to minimize any priming or experimenter demand effects before eliciting the main discrimination outcome, so only post-discussion choices are used in this analysis. Baseline and endline attitudes are likely to be positively correlated under a reasonable assumption: that pro-participants are not *so* persuasive that they persuade others to be even more pro-trans than themselves. Nevertheless, the lack of baseline measures is an important limitation to this analysis.

[9] Since being in a group with others with a persuasive personality does not predict greater reductions in discrimination, this indicates that it is the *arguments* rather than the *people* that are more persuasive.

discriminatory behavior) and the prescriptive norm (widespread agreement that discrimination is wrong). Control group participants discriminated substantially, yet 93% described discrimination as wrong in response to vignettes. This disconnect can plausibly explain the three sources of asymmetry that could explain the shift away from discrimination: (i) pro-trans participants speaking up more because there is minimal social harm from advocating for transgender workers, and more to gain from persuading others; (ii) participants more readily updating their perception of the descriptive norm in favor of pro-trans behavior to align more closely with the prescriptive norm; and (iii) participants being particularly responsive to morally-framed arguments in favor of transgender workers because they align with their existing perceptions of what is "right". These suggestive patterns, though not conclusive, point to promising contexts for horizontal communication interventions where similar norm disconnects exist. They simultaneously indicate that horizontal communication is not a silver bullet, and may be actively harmful where people subscribe to the view that discrimination is morally correct.

The key contribution of the paper is to show that discrimination can be rapidly reduced by generating horizontal communication about a minority within an existing social network. While many previous studies have shown that social contact between in-groups and out-groups can affect discrimination (**???????????**), I show that even communication between in-group members about an out-group can reduce discrimination.

The most closely related studies are **?** and **??**, which show that trained door-to-door and phone canvassers in the US can reduce prejudice through short conversations, even when the canvasser is not from the minority group. My study differs in three main ways. I focus on communication *within an existing social network* – among neighbors who know each other well, rather than canvassers talking to strangers. Understanding whether existing social networks can generate norm change raises new conceptual issues: network members may exert stronger influence on each other, but also face higher costs for challenging discriminatory norms or be too homogeneous to generate changes in behavior. Second, I examine *undirected* communication, where participants have no explicit incentive or instruction to reduce discrimination, unlike trained canvassers with clear directives to decrease bias. The endogenous choice to speak up against discrimination appears to be important in my context and reveals how norms might shift organically without external intervention in other contexts. Third, I examine communication that includes making revealed-preference choices, with revealed-preference discrimination outcomes. This reduces the possibility of cheap talk both during the intervention and when measuring the outcome.

Other papers evaluate the effect of ingroup contact on outgroup prejudice. **?** find that ingroup contact backfires, and other papers typically find null results (see **?**). I contribute to this scarce literature by documenting a large reduction in discrimination that contrasts with previous studies, and by focusing on semi-structured communication about a minority among the ingroup, rather than unstructured ingroup contact that is often bundled with other treatments (e.g., skill training).

The discussion generates a misperceived norm that reduces discrimination, in line with work

on pluralistic ignorance (misperceptions about norms) and how they are affected by communication (**???????**). My work differs from previous papers in three ways. First, participants generate the norm shift themselves through horizontal communication, rather than in reaction to statistics about norms delivered top-down by the experimenter. My results imply that more naturalistic communication could rapidly shift behavior, even without external intervention. Second, I expand the literature on persuasion by documenting the importance of moral arguments and the endogenous choice to persuade (**??**), highlighting the potential role of moral "activists" in driving social change. Third, while other work implies that social change occurs only in the presence of pluralistic ignorance, my results suggest a different precondition: when there is a wedge between what people do and what they believe is right. If people discriminate even when they think it is wrong, horizontal communication is likely to be especially helpful because pro-minority communicators will be more willing to speak up, more able to credibly signal an anti-discriminatory norm, and more able to use persuasive moral arguments.

Finally, I examine potential policy levers for reducing discrimination against LGBTQ+ persons in a lower- or middle-income country (LMIC). Even though such discrimination may have significant costs (**??**), very little research in LMICs has examined its effects and causes (**?**).[10] I contribute by evaluating the potential to reduce this discrimination using horizontal communication.

2 Context: Transgender community in India

This study examines discrimination against a historically marginalized community in South Asia largely composed of transgender women, who in the state of Tamil Nadu are called *thirunangai*.[11] This group has a longstanding cultural and religious role in Indian society (**??**). Their visually recognizable identity, however, leaves them particularly susceptible to discrimination (**??**). Economic discrimination against this group is multi-faceted. Transgender people are often excluded from traditional forms of paid employment, pushing many into poverty and sex work (**????**), resulting in low levels of formal employment (**?**). There are likely at least 0.5–1.25 million transgender people in India (**???**), implying that such discrimination results in large welfare and efficiency costs (**?**).

Widespread prejudice and discrimination appear to be increasingly at odds with social norms that penalize discriminators or promote pro-trans activism. Discrimination, though common, can generate social disapproval: in the study control group, discriminatory scenarios were rated as "wrong" 93% of the time. This aligns with other survey evidence that indicates widespread support for protecting transgender people from discrimination (**?**). The context is thus analogous to other settings where private prejudice is relatively common, but its expression may be inhibited by social sanctions for prejudiced behavior (**??**).

Legal changes may have contributed to the decreasing acceptability of discrimination. In 2014,

[10] There are notable exceptions. **?** shows that explaining to Ugandan citizens that homosexuality is legal in other countries leads to a backlash effect, worsening participants' opinions of those countries. **?** show that Latin American soap operas with LGBTQ+ characters also generate backlash, and **?** show evidence of housing discrimination for some LGBTQ+ persons. Research on anti-LGBTQ+ discrimination in Europe and the US has examined its magnitude and nature (**??????????**), along with strategies for reducing it (**?????**).

[11] Throughout the paper, for simplicity, I refer to people from this community as "transgender people".

the Supreme Court recognized all constitutional rights for transgender persons, along with their right to identify as a third gender, and encouraged government initiatives to combat anti-transgender stigma. But awareness of these recent changes remains low. 36% of control participants either believe that trans individuals do not have any legal status in India, or cannot identify a single legal right they hold.

The study took place in Chennai, the largest city in the state of Tamil Nadu. The urban setting is advantageous because (i) approximately 80% of thirunangai live in urban areas (**?**), so urban residents are aware of and can visually recognize transgender people; and (ii) urban residents are familiar with online delivery services, allowing me to use delivery service market research for the study.[12]

3 Experimental design

3.1 Design overview

3397 participants in Chennai, India took part in the field experiment. The experiment measures the effect of horizontal communication (group discussions) on subsequent hiring discrimination against transgender workers.

The main goals of the experimental design were: (i) measuring discrimination in choices with real stakes; (ii) generating horizontal communication *between* participants about transgender people, without the purpose of the study being obvious; (iii) understanding the mechanisms driving the effects; and (iv) benchmarking effects against the effect of delivering information about transgender persons' legal rights.

All treatments and primary data collection occurred in one session that lasted approximately 1 hour. To allow for a group discussion, 3 enumerators recruited and interviewed 3 respondents (a "group") at the same time.

To measure hiring discrimination, participants were offered a free grocery delivery to their home, and made 10 binary choices between different delivery workers and grocery bundles. One choice was randomly selected to be implemented: between 2 and 9 weeks later, the selected worker delivered the selected grocery bundle, and participants were asked follow-up questions.

The first four choices served as treatment variation while the final six measured outcomes. During the *treatment round*, some participants discussed their preferences in groups, while others chose privately. Then during the *outcome round*, all participants made their final six choices individually and privately. These private choices are the main outcome, allowing me to measure how discussions affected *private* behavior after the discussion.

The main treatments examined how horizontal communication affects discrimination, and why. The main effect of interest compares (i) *3-person discussion*, in which all 3 participants had a discussion about their preferred hiring options and made collective choices; and (ii) *No*

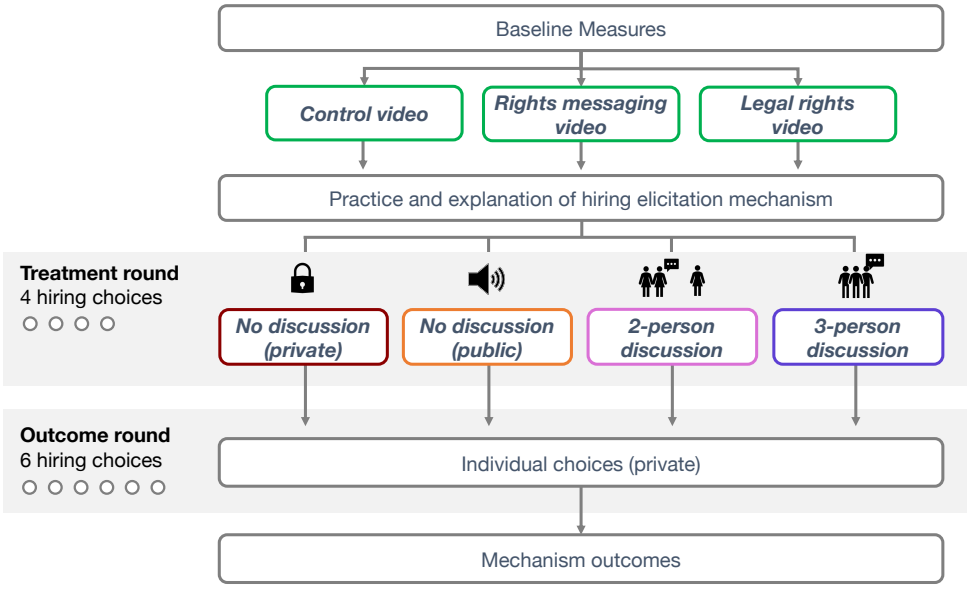
[12] 80% of the sample say that they have previously ordered goods to be delivered to their home using an app, reflecting the popularity of meal delivery services such as Swiggy and Zomato. The market research framing is also not so unusual in this context: 29% of the sample have previously taken part in a market research survey or received a free item as a promotion.

discussion (private), a control condition in which all participants made private individual hiring choices even during the treatment round. Two further treatments explore the mechanisms: (iii) *2-person discussion*, in which 2 participants had a discussion and made a collective choice, while the third participant silently listened; and (iv) *No discussion (public)*, in which all participants made individual hiring choices that they knew would later be revealed to the other group members.

A secondary source of treatment variation tested the effects of top-down communication about legal rights. I cross-randomized a video shown to participants before they made any hiring choices. Participants either saw (i) a *legal rights* video containing information about the legal rights of transgender people, (ii) a *rights messaging* video containing persuasive messaging in favor of transgender rights, or (iii) a *control* video that did not mention transgender rights.

?? shows a summary of the experimental design for the main session (**??** gives further detail). I describe the design further below, with additional details in Appendix **??**.

Figure 1: *Summary of experimental design*



Notes: More detail is given in Appendix Figure **??**.

3.2 Sample and recruitment

Participants were recruited from urban areas in Chennai between March and July 2023 through direct household canvassing and introductions from community leaders (see **??** for survey locations). All participants were offered a free grocery delivery as an incentive to participate. Participants were aged 20-65 and could read Tamil, and the median per capita food expenditure (Rs. 2000 per month; 87.40 USD PPP) was similar to a representative 2012 sample of urban Tamil Nadu residents (**?**).

We framed the study as market research about grocery delivery preferences to minimize experimenter demand effects. Participants were truthfully told that we were trying to understand people's preferences for grocery delivery options. Only 8% of participants guessed the study's

Figure 2: *Example of one of the binary choices participants face*

focus on transgender workers after completing their main hiring choices. These participants do not drive the treatment effects (see Appendix **??**).

To allow for the group discussion, 3 enumerators always recruited and interviewed 3 respondents at the same time (including for the control group). To avoid recruitment strategies that differed across treatments, enumerators were blind to treatment status before starting the survey.

Groups consisted of neighbors or acquaintances from the same street or locality to facilitate natural interactions. Group members knew each other 98% of the time, described each other as family or friends 41% of the time, and as neighbors 62% of the time. To avoid hierarchical relationships in which one group member dominated a discussion, we recruited either all-male or all-female groups without multiple members of the same household. The sample was predominantly female (85%). The study's focus on deliveries was more relevant for females, since they were more likely to be responsible for managing food expenditures and receiving deliveries (88%) than men (59%).

3.3 Hiring choices

Participants made 10 binary choices between delivery options, with one randomly selected to be implemented. Each option showed the workers' photos and the grocery items offered (see Figure **??**), sometimes inducing a trade-off between a preferred worker and preferred items.

In each choice, there was a "*benchmark*" option (cisgender male), and an "*alternative*" option (either cisgender male, cisgender female, or transgender).[13] Throughout the paper, I measure anti-transgender discrimination as the reduction in probability of choosing the alternative when the alternative was transgender.

4 of the 20 workers were transgender. This proportion was chosen to ensure sufficient power without making the experiment's purpose too obvious. We randomized the alternative worker's position (left or right), the order of choices, and the specific photos within gender

[13]This reduces the number of gender combinations, thus increasing power on the male-to-trans comparison, although it does not allow a direct comparison of trans and female workers.

categories. We never showed the same worker twice to a participant.[14]

Participants could visually identify the transgender workers in the photos. A validation study in late 2022 (N=114) showed that participants correctly identified transgender photos as being transgender 97% of the time (Appendix Table **??**).

I measure participants' willingness to pay to avoid certain workers by randomly varying the items offered in each choice, so in 40% of cases one worker offered more items than the other. Each worker offered either 1 item (masala spice mix), 2 items (masala and tea), or 3 items (masala, tea, and ghee). The randomization was balanced across genders. The clear ranking of the bundles made the tradeoff between item value and worker characteristics clear for participants. The value of the item bundles was substantial relative to participants' consumption, corresponding to 103%, 234%, and 365% of median daily per capita food expenditure.

To ensure participants anticipated some social contact with the worker, participants learned they would have a 15-minute conversation to discuss their satisfaction with the service with their chosen worker during the delivery. When selecting workers, participants were told to consider the worker's characteristics, the items they offered, and this 15-minute conversation.

To test if discrimination was statistical (e.g., stemmed from beliefs about worker reliability), some choices included truthful extra signals of worker quality. These included a "reliability score" (the proportion of successful deliveries from timed training exercises), work experience, and languages spoken (see Appendix **??**). These were sampled to be balanced across genders.

Implementing choices. To ensure incentive-compatibility, enumerators used scratch cards to randomly select 1 of the participants' 10 choices to be implemented. The participant received the chosen items from the chosen worker 2–9 weeks later, and the worker carried out a 15-minute follow-up survey at the same time. To minimize risk to transgender workers, the randomization selected choices involving transgender workers in fewer than 1% of cases.[15]

We helped participants understand the randomization scheme using a practice round. Mimicking the main hiring choices, participants made 4 binary choices between items worth under Rs. 5, with enumerators using scratch cards to implement one choice. We asked comprehension checks before the practice and main hiring round. The high success rates (92% correct first time for the practice round, 86% for the main round) indicated good understanding.

[14] Photos were selected from a pool of 20 males, 21 females, and 13 transgender people. Cisgender photos were of survey enumerators who carried out the main survey. Participants were never shown pictures of the enumerator-team that interviewed them to avoid response bias. Transgender photos showed real enumerators we recruited who had agreed to carry out deliveries if selected, but did not carry out the main survey. Later mechanism outcomes (that did not require workers to deliver goods) used stock headshot photos.

[15] This design protected transgender workers from potential abuse or violence while maintaining incentive-compatibility: participants could receive a delivery from any chosen worker, but the probability was lower for transgender workers. When transgender workers did make deliveries, they were accompanied by a team of 2-3 enumerators and a supervisor, and minimized participant interaction. See Appendix **??** for discussion of the ethical considerations.

3.4 Treatments

3.4.1 Discussion arms

The first four hiring choices (the *treatment round*) varied how participants communicated about their decisions. Two of these four choices included a transgender worker. Groups were randomly assigned to one of four conditions, with randomization stratified by participant gender and survey team:

1. **3-person discussion** (N=890). 3 respondents discussed their delivery and worker preferences, and then made *joint* choices (see more detail in Section **??**).
2. **2-person discussion** (N=549). Two randomly selected "*speakers*" discussed and chose together, while the third (the "*listener*") observed silently and listened to the speakers' choices and justifications. This mechanism arm tested whether the effects were driven by interpersonal influence, or whether active participation was necessary.
3. **No discussion (public)** (N=599). Participants made silent individual choices, knowing that their choices would be later announced to others in their group. This mechanism arm exogenously increases social image concerns, and evaluates whether this alone affected participants' choices. I also varied the timing of the announcement to test whether observing others selecting a transgender worker could affect subsequent discrimination, and how this compared to listening to a discussion. 2 randomly selected "*observers*" learned others' choices before making their private outcome-round decisions, while the 3rd ("*non-observer*") learned others' choices only afterward.[16] These observers were designed to test whether changes in the perceived group norm of discrimination were sufficient to affect discrimination.
4. **No discussion (private)** (N=1365). Participants made choices individually and privately, serving as the control condition.

Enumerator observations indicate that participants correctly followed the protocols.

3.4.2 Design of the discussion

Treatment round activities for the three group conditions occurred with all participants together, usually in someone's home or a nearby common area like a building courtyard. These group sessions lasted about 10 minutes. In contrast, *No discussion (private)* participants completed their interviews individually and out of earshot from one another.

Both discussion treatments required participants to make collective decisions for each pair. When these collective choices were selected by the scratch cards, each group member separately received the same items from the same worker. The collective choices created an incentive to actively participate in the discussion by ensuring the discussion had real stakes for participants. In contrast, participants in *No discussion* treatments made individual choices.

Participants freely discussed their preferences and worked to resolve disagreements about options before coming to a conclusion about which option to select. I document the types of

[16] Participants were not told about the distinction between *observers* and *non-observers* until after the end of the treatment round, in order to avoid this affecting treatment round choices.

statements made during the discussion in Section **??**. The enumerator leading the discussion never mentioned the word *transgender* themselves to minimize demand effects and to measure only the effects of horizontal communication arising between participants (see Appendix **??** for the discussion script and further details).

3.4.3 Rights videos

I cross-randomized a second set of treatments to compare to the effect of top-down communication about minority rights. Participants viewed one of three 80-90 second videos about rights before making their hiring choices. The video was narrated in Tamil by a local research team member (who was not shown). While most content remained constant across videos and explained consumer and worker rights for delivery services, the videos differed in one key example used to illustrate rights (Appendix **??** contains full scripts and further details):

1. **Legal rights video** (N=1135). This informed participants that transgender people have legally protected rights in India: *"The Supreme Court of India, the most powerful legal institution in the country, gave transgender people all the same fundamental rights as others under the Constitution of India. The law therefore gives them the right to housing, employment, and education without discrimination. All these rights that you have, they also have according to the law."*

2. **Rights messaging video** (N=1135). This argued that transgender people *should* have rights but did not mention legal protections, testing whether rights-based narratives could affect discrimination without institutional backing: *"Transgender people should have the same fundamental rights as others in India. They should have the right to housing, employment, and education without discrimination. All these rights that you have, they should also have."*

3. **Control video** (N=1135). Participants received neutral information about voting rights: *"Some people have the right to vote. If you have the right to vote, you can elect your representatives. That means you can choose who should be in power and who should make decisions on your behalf."*

To maximize understanding, participants were given comprehension checks and corrected if needed.

3.5 Data collection phases and samples

Figure 3: *Sample sizes and timeline*

	Phase 1 March – April 2023	Phase 2 May – July 2023	Total
No discussion (private)	N = 603	N = 756	**N = 1359**
No discussion (public)		N = 599	**N = 599**
2-person discussion		N = 549	**N = 549**
3-person discussion	N = 576	N = 314	**N = 890**

Data collection occurred in two phases (see **??**). Phase 1 (March–April 2023) focused on the

main treatment effect, comparing private choices to 3-person discussions. Phase 2 (May–July 2023) added the *No discussion (public)* and *2-person discussion* arms after receiving additional funding in order to understand the mechanisms behind the effects of the discussion.

When analyzing the data, I primarily use three different samples: (i) the *3-person discussion sample* which measures the effect of the 3-person discussion by including *No discussion (private)* and the *3-person discussion* arms in both phase 1 and 2; (ii) the *phase 2 sample* which uses only phase 2 and includes all treatment arms to analyze the effect of the mechanism treatments; and (iii) the *video sample* that includes all data from all phases since the videos were cross-randomized across all discussion arms.

3.6 Pre-analysis plan

I preregistered the design of both phases, and document deviations from the pre-analysis plan in Appendix **??**. I changed the main specification to exclude interaction terms between discussion and video treatments to ease interpretation and increase power (I also show interacted specifications in **??**). The other main deviations were driven by unexpectedly low survey productivity in phase 1, which tightened budget constraints. These include: (i) dropping a mixed-video arm in which each group member saw different rights videos; (ii) delivering groceries after 2–9 weeks instead of 1 week; and (iii) collecting fewer mechanism outcomes.

3.7 Balance checks

The treatment groups were well balanced on key characteristics (Tables **??**–**??**), with joint F-tests showing no systematic differences from the control group for any treatment. As expected, given the large number of comparisons, individual variables show some statistically significant differences across treatment groups. *3-person discussion* participants were more likely to have employed someone in the last 2 years (**??**), and *rights messaging video* participants came from slightly larger households, with a slightly lower per capita food expenditure (**??**). I use LASSO to select all controls that predict both treatment status and outcomes (**?**; see Appendix **??**), so these imbalances are unlikely to affect my results.

3.8 Outcome and specification

The pre-specified primary outcome is participants' individual choices in the outcome round of hiring. The design thereby estimates the causal effect of the discussion on participants' *post-discussion* private discrimination. The outcome round occurred after group activities had finished, with 94% of respondents confirming that other group members could not hear their responses. While these choices were designed to be as private as possible, minimizing social image concerns, participants' choices may still have been affected by social image concerns if (i) they knew each other and would discuss choices later; (ii) they anticipated that neighbors would observe the delivery worker when the delivery took place; and (iii) enumerators observed the answers given by respondents. While I cannot rule out channel (i), for robustness I use an "extra private" outcome that addresses channels (ii) and (iii) (see Appendix **??**).

The outcome round included 6 binary choices, two of which included a transgender worker. The main specification for participant i in group j, making a choice for the pair of workers k, is:

$$\begin{aligned} ChooseAlternative_{ijk} = \sum_{\tau \in \mathcal{T}} \beta_\tau \left(Treat_{\tau ij} \times Trans_{ijk} \right) + \gamma Trans_{ijk} + \sum_{\tau \in \mathcal{T}} \delta_\tau Treat_{\tau ij} \\ + \mathbf{X}'_{ijk}\Gamma_0 + (\mathbf{X}'_{ijk}\Gamma_1 \times Trans_{ijk}) + \varepsilon_{ijk} \end{aligned} \tag{1}$$

$ChooseAlternative_{ijk} = 1$ if i selects the *alternative* worker in pair k (who could be transgender or non-transgender), and is 0 when i selects the male *benchmark* worker. $Trans_{ijk} = 1$ if the alternative worker is transgender, and is 0 if the alternative worker is non-transgender (cisgender male or female). The alternative worker is always compared to a male benchmark worker. $Treat_{\tau ij}$ is a dummy for whether i is in treatment arm $\tau \in \mathcal{T}$, where the set of treatments $\mathcal{T}$ is either (i) a dummy for the 3-person discussion, (ii) dummies for each discussion-arm treatment, or (iii) dummies for each rights video.

$\mathbf{X}_{ijk}$ is a vector of controls that are included in some specifications. Controls are interacted with $Trans_{ijk}$ to control for differences in discrimination driven by observables. The controls include stratum fixed effects, differences in items offered, differences in reliability score, the benchmark worker's reliability score, whether the reliability score was shown, question order fixed effects, a dummy for whether the alternative worker was shown on the right, and data collection phase fixed effects. When analyzing the discussion-arm treatments, I control for the rights videos, and vice versa. I use double LASSO (**?**) to select additional controls that predict both the treatment and outcome variables (see Appendix **??**).

Throughout the paper, I define discrimination as the reduction in the probability that a worker is chosen because they are transgender (relative to non-transgender), conditional on other characteristics of the delivery options, such as the items on offer.

The main treatment effects are thus given by the coefficients β_τ, which describe the reduction in discrimination caused by the treatments. When interacted controls are not included, γ describes the control group's discrimination against transgender workers. Standard errors are clustered at the group-of-3 level. For tables in the main text, I use randomization inference to calculate p-values (**?**). Since I have only one primary outcome, I do not correct it for multiple hypothesis testing.

3.9 Mechanism outcomes

The experiment also measured several mechanisms (see Appendix **??** and the relevant results section for more detail on each).

Baseline measures included questions on (i) demographics; (ii) social desirability bias based on **?** (phase 1 only); (iii) proximity of relationships between group members (phase 2 only); and (iv) a persuasiveness index to measure how persuasive an individual was likely to be in a discussion (phase 2 only). These questions were intermingled with questions about deliveries

to reinforce the study's framing as market research.[17]

Treatment round choices. Hiring choices during the *treatment round* are a pre-specified secondary outcome. These allow me to examine what choices were made *during* (rather than *after*) the discussion.

Group observations. During the group activities, an enumerator who was not facilitating the discussion noted observations about the group activity (e.g., who spoke first, who spoke in favor of a transgender worker). We recorded and transcribed the audio of the 3-person discussions,[18], and translated the Tamil transcriptions to English, yielding a dataset at the (sentence $\times$ group $\times$ choice) level. The enumerator observations and audio transcripts are complementary since it was not possible to determine *who* said what in the transcripts.

Post-hiring mechanisms. Immediately after the hiring choices, we elicited further mechanism outcomes. Pre-specified secondary outcomes were: (i) predictions about the private hiring choices of other unknown people in the study; (ii) predictions about the private hiring choices of other participants in the same group; (iii) self-reported disapproval of discrimination when presented with discriminatory scenarios; (iv) a double list experiment (**??**) measuring the proportion of people agreeing with a discriminatory statement; and (v) questions about the legal status of transgender people. Additional exploratory mechanisms included: (vi) beliefs about reliability (the probability that the worker will complete a delivery); (vii) hiring choices for a private grocery pick-up unobservable to neighbors (phase 2 only); (viii) a measure of salience of the word "transgender" using a surprise recall task; (ix) two measures of participants' beliefs about the purpose of the study; and (x) self-reported reasons for their hiring choices (e.g., the most important factors when making their decision).

Follow-up survey. When the delivery occurred, an average of 35 days after the initial survey (SD: 14 days), we conducted a short (15-minute) survey to measure the persistence of treatment effects. As a pre-specified secondary outcome, we asked 6 more hypothetical hiring choices with new worker photos and different grocery items. We made it clear to respondents that their choices would not result in actual deliveries.

As pre-specified, I correct for multiple hypothesis testing within sets of secondary outcomes, namely for attitudes (the list experiment and discrimination disapproval measure), and for norms (the predicted choices for community and own group).

[17] I did not include baseline measures of anti-transgender attitudes or pre-treatment hiring choices. While this would have increased power and yielded insights into the relationship between baseline attitudes and group discussions, it also risked undermining the credibility of the main results for two reasons. First, asking additional questions about transgender people risked making the true purpose of the study more salient, exacerbating concerns of experimenter demand effects. This concern was especially severe for attitude questions that explicitly talk about discrimination against transgender people, contrasting with the hiring questions that are subtler and less obviously focused on discrimination. Second, evidence suggests that people prefer to act consistently with previous actions (**?**), and persuade themselves to ensure their preferences align with their previous actions (**?**). If true, measuring baseline discrimination would anchor behavior to a pre-treatment state, and lead treatment effects to be underestimated.

[18] Consent for audio recordings was refused in 16% of discussions. The discussion's effects are not significantly different for groups that refused to be recorded (p of difference = 0.84).

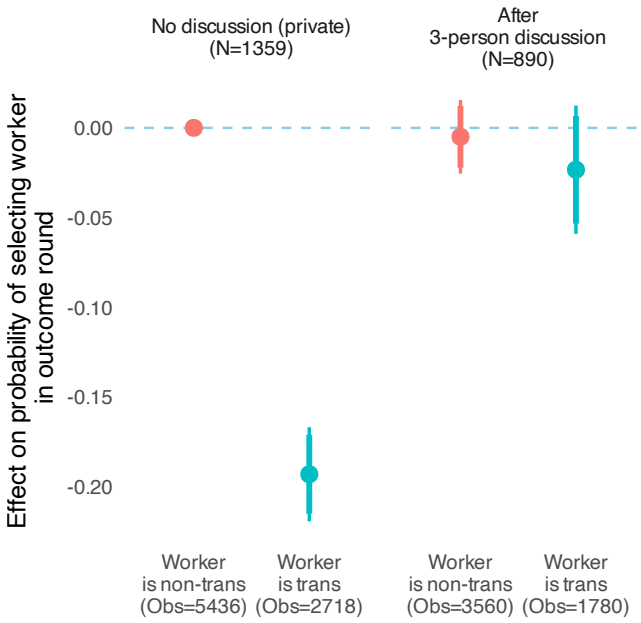
4 Results

4.1 Effect of 3-person discussion

The 3-person discussion leads to large reductions in discrimination in the private choices made *after* the discussion in the later outcome round (Table **??** and Figure **??**). In the control group, *No discussion (private)*, there is substantial discrimination: participants are 19 p.p. less likely to select a transgender worker than a non-transgender worker ($p<0.001$, Table **??**, Column 1). But if participants were earlier involved in a group discussion and collective hiring decision, the probability that they chose a transgender candidate in their individual choices increases by 17 p.p. ($p<0.001$). Participants in the discussion arm thus do not discriminate against transgender workers on average ($p=0.30$).

The treatment effect of the 3-person discussion is robust to the inclusion of controls (Column 2), to including only choices that involve a transgender worker (Column 3), and to restricting the sample to participants who did not see a video discussing transgender rights (**??**).[19]

Figure 4: *Effect of 3-person discussion on private choices in outcome round*



Notes: This graphically shows the main results from Table **??**. The y-axis shows the effect on the probability of choosing the *alternative* worker relative to the case where *Worker is non-trans* in the *No discussion (private)* arm. 90% and 95% confidence intervals are based on column (1) of Table **??**, based on standard errors clustered at the group-of-3 level.

To benchmark the size of the reduction in discrimination, I use the random variation in grocery items to infer the willingness to pay to avoid a transgender worker (**??**). In the *No discussion (private)* arm, participants are on average willing to sacrifice items worth Rs. 127 (approx. 6.30 USD PPP) to avoid selecting a transgender worker, corresponding to 1.9x the median daily per capita food expenditure in the sample. By contrast, in the *3-person discussion* arm, the

[19] Cross-randomized designs that do not account for interaction terms can yield incorrect inference (**?**). I show interaction terms for completeness (**??**), but since the results hold for participants who only saw the control video, interaction effects cannot be driving the main effects of the discussion.

Table 1: *Effect of 3-person discussion on private choices in outcome round (3-person discussion sample, Phases 1 and 2)*

	Chose worker in private outcome round (=1)		Chose trans in private outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans × 3-person discussion	0.175***	0.168***	
	(0.022) [<0.001]	(2.2×10^{-2}) [<0.001]	
Worker is trans	−0.193***		
	(0.013) [<0.001]		
3-person discussion	−0.004	0.001	0.167***
	(0.011) [0.500]	(1.0×10^{-2}) [1.000]	(0.020) [<0.001]
Relative # items offered		0.128***	0.097***
		(5.0×10^{-3}) [<0.001]	(0.008) [<0.001]
Relative reliability score		0.017***	0.013***
		(3.0×10^{-3}) [<0.001]	(0.005) [<0.001]
Reliability score is shown (=1)		0.022***	0.040***
		(8.0×10^{-3}) [<0.001]	(0.012) [<0.001]
Num. observations	13 494	13 494	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Outcome mean†: no discuss (priv.), worker non-trans	0.61	0.61	
Outcome mean†: no discuss (priv.), worker trans	0.42	0.42	0.42
Controls		X	X
Controls interacted with worker is trans		X	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. Column (3) only includes choices that involved a transgender worker. In columns (1) and (2), the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was selected in the private choices in the *outcome round*. In column (3), it is whether the transgender worker was selected. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. The specification used is seen in equation **??**. Controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; and the controls selected by double LASSO (see Section **??**). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown. Relative # items offered is the number of items offered by the *alternative* worker minus the number of items offered by the male benchmark worker. Relative reliability score is the reliability score (out of 10) of the alternative worker minus the benchmark worker. *Reliability score is shown* is 1 when the reliability score is shown. Relative reliability score is coded as 0 when it is not shown.

† The dependent variable mean when the worker is trans in the *No discussion (private)* arm indicates that the transgender worker was selected (rather than the male benchmark worker) 42% of the time. The mean when the worker is male or female in the *No discussion (private)* arm is above 50% because participants on average prefer female alternative workers to the male benchmark workers.

willingness to pay to avoid is Rs. 13 (p of difference < 0.001), and is no longer significantly different from 0 (p=0.265). The results therefore suggest that the discussion generates a large short-run reduction in discrimination.

The effect size is similar when examining only *costly* discrimination, i.e., when participants avoid a transgender worker who offers more items, has a higher reliability score, or both (**??**, column 1).[20] In the *No discussion (private)* arm, even when shown a transgender worker that dominates on items or reliability score, or both, participants still select the non-transgender worker 47% of the time. By contrast, in the *3-person discussion* arm, this figure has reduced to 30% (difference: 17.5 p.p., $p<0.001$). The discussion also increases the probability of selecting a transgender worker when they offer fewer items or have a worse reliability score (12.7 p.p., $p<0.001$), implying that it also increases *positive* discrimination in favor of transgender workers.

There is heterogeneity in levels when separating the analysis of non-transgender workers into males and females (**??**).[21] Appendix **??** shows further heterogeneity results, indicating that (i) participants trade off their preferences between items and worker; (ii) participants statistically discriminate against transgender workers, but the discussions' effects are not driven by this; (iii) the discussion reduces discrimination by the same amount for male and female participants.

4.2 Effect of transgender rights videos

The videos about transgender rights reduce discrimination by significantly less than the discussion (**??**). Both the *Rights messaging* video and the *Legal rights* video significantly increased the probability of selecting a transgender worker in the outcome round (β=5.8 p.p. and 10.3 p.p., $p<0.01$). There is some evidence that the legal rights video has a stronger effect than the rights messaging video ($p \in [0.00, 0.00]$, depending on the specification). Endorsing transgender rights thus appears to reduce discrimination somewhat more effectively when it is backed by the legal authority of the Supreme Court. This implies that the law can be an important tool for reducing societal discrimination, and that raising awareness of the legal rights of minorities may be an underrated policy lever for addressing discrimination. However, the effect of top-down communication about the law is only 59% as large as the effect of the group discussion (p of difference $\in [0.002, 0.04]$). Appendix **??** shows further results on the rights videos, including (i) a manipulation check showing that the *legal rights videos* affect people's beliefs about transgender rights; (ii) the interactions between the rights videos and the discussion; and (iii) how the rights videos affected the mechanism outcomes.

[20] I do not find evidence of an implicit preference against transgender workers, in which they are discriminated against mostly when there is no plausible deniability (e.g., when other attributes are different) (**?**). Transgender workers are not disproportionately favored when their attributes are the same as the alternative worker, compared to when they are dominated or dominating (**??**).

[21] Female workers were the most preferred gender in both treatment conditions and were selected 72% of the time over the *benchmark choice* (who was always male). Male workers, always being compared to other males, were mechanically selected around 50% of the time. Transgender workers, however, were selected 42% of the time in the *No discussion (private)* arm, but 59% in the *3-person discussion* arm. This implies that males were preferred to transgender people in the control condition ($p<0.001$), but transgender people were preferred to males in the treatment condition ($p<0.001$).

Table 2: *Effect of rights videos on private choices in outcome round*

	Chose worker in private outcome round (=1)		Chose trans in private outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.175***		
	(0.016) [<0.001]		
Rights messaging video	−0.013	−0.016***	0.053***
	(0.011) [1.000]	(1.0×10^{-2}) [<0.001]	(1.9×10^{-2}) [<0.001]
Legal rights video	−0.019	−0.021***	0.081***
	(0.011) [0.500]	(1.0×10^{-2}) [<0.001]	(1.9×10^{-2}) [<0.001]
Worker is trans × Rights messaging video	0.058***	0.070***	
	(0.023) [<0.001]	(2.1×10^{-2}) [<0.001]	
Worker is trans × Legal rights video	0.103***	0.104***	
	(0.022) [<0.001]	(2.0×10^{-2}) [<0.001]	
Num. observations	20 382	20 382	6794
Num. participants	3397	3397	3397
Num. groups	1134	1134	1134
Outcome mean: control vid., worker non-trans	0.62	0.62	
Outcome mean: control vid., worker trans	0.45	0.45	0.45
Controls		X	X
Controls interacted with worker is trans		X	
p(Rights messaging video=Legal rights video)	0.000	0.000	0.000
p(Legal rights video=3-person discussion)	0.024	0.037	0.002

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all participants in both phases, in all discussion-arm treatments. Controls include dummies for the discussion-arm treatments. The specifications are otherwise the same as **??**.

4.3 Persistence after 2–9 weeks

To examine whether the effects of the discussion and the rights videos persist over the medium-run, I ran a follow-up survey when the delivery was carried out. This survey took place an average of 35 days after the initial survey (SD: 14 days). 95.7% of the sample were found, and there was no evidence of differential attrition (**??**). For the follow-up, discrimination was measured using 6 hypothetical hiring choices designed to be as similar as possible to the main hiring choices. All of these questions were asked individually and in private. The questions used a new set of worker photos and new types of grocery items.

The 3-person discussion led to reductions in discrimination in these hypothetical choices that were still present after 2-9 weeks (**??**). Participants were approximately 4 p.p. more likely to select transgender workers in the hypothetical follow-up choices (p=<0.001). Approximately 25% of the short-run effect thus remained after around 1 month. This is comparable to persuasion decay rates seen in the political science literature (e.g., **?**, who estimate that 10-15% of the initial effects of US TV political advertisements on voting remain after 30 days), although is less persistent than the effects in **?** and **?**. By contrast, the videos about transgender rights did not lead to a detectable persistent effect on discrimination ($p \in [0.00, 0.50]$, **??**).

Since the follow-up choices are hypothetical, the results are more vulnerable to concerns about experimenter demand effects and social desirability bias. However, experimenter demand effects are likely to be substantially *larger* for the rights videos (which increase the probability of successfully detecting the purpose of the experiment from 5% to 10%, p<0.001, see Appendix

Table 3: *Medium-run effects of discussions on hypothetical hiring choices (2-9 weeks)*

	Chose worker in follow-up round (=1)		Chose trans in follow-up round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.127***		
	(0.013) [<0.001]		
3-person discussion	−0.004	−0.005***	0.043***
	(0.011) [1.000]	(0.011) [<0.001]	(0.019) [<0.001]
Worker is trans × 3-person discussion	0.054***	0.048***	
	(0.021) [<0.001]	(0.021) [<0.001]	
Num. observations	12 780	12 780	4254
Num. participants	2130	2130	2127
Num. groups	745	745	745
Outcome mean†: no discuss (priv.), worker non-trans	0.62	0.62	
Outcome mean: no discuss (priv.), worker trans	0.49	0.49	0.49
Controls		X	X
Controls interacted with worker is trans		X	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phases 1 and 2. Controls include dummies for the rights videos, as well as the other controls specified in Table **??**. In the follow-up survey, workers in a pair always had the same reliability score and offered the same number of items. Specification is otherwise the same as Table **??**. As prespecified, for analysis, I drop the 0.4% of the sample who were randomly selected to actually receive a delivery from a transgender worker.

??). The fact that there were no persistent impacts for rights videos thus suggests that the discussions' effects were not driven by experimenter demand. The results therefore suggest that the large short-term effects of the discussion on discrimination may translate into medium-run effects. This raises the possibility that even short interventions involving horizontal within-group communication could have persistent effects on behavior.

5 Mechanisms: how does the discussion affect post-discussion discrimination?

How did group discussions reduce discrimination? I address this question in two sections. In this section, I use mechanism treatments and outcomes to identify how within-discussion behavior affected post-discussion choices at all. I show experimental evidence in favor of inter-personal persuasion, and suggestive correlational evidence that participants persuaded each other primarily by influencing perceived norms around hiring transgender workers, rather than by changing attitudes or beliefs. I then investigate why discussions created a large mean shift away from discrimination (Section **??**). Since standard models of communication (e.g., based on rational Bayesian agents) typically predict no changes in average beliefs, the substantial treatment effects require identifying asymmetries in the communication process. I provide suggestive evidence of three such asymmetries: pro-transgender participants spoke up more frequently, participants incorrectly updated and misperceived the strength of anti-discriminatory norms, and pro-trans arguments based on morality appear to have been particularly persuasive.

Appendix **??** also provides evidence against a number of alternative explanations of the results, including (i) other characteristics of the photos of workers, such as perceived caste; (ii) social image concerns that affect even private, post-discussion choices; (iii) increased contemplation about choices; (iv) experimenter demand effects or social desirability bias; (v) increased salience

of the notion of transgender; (vi) cheap talk due to low stakes; and (vii) discussion facilitators influencing discussions to be pro-trans.

5.1 Candidate mechanisms for affecting post-discussion behavior

First, I explore three candidate mechanisms that could explain how behavior during the discussion could affect private post-discussion discrimination:

(1) *Norm-based persuasion*. Participants could persuade each other to discriminate less by changing perceived social norms. After observing frequent pro-trans statements and choices, participants may come to view discrimination as less socially acceptable or less prevalent. This perceived norm could then influence their post-discussion choices if they are motivated to adhere to social norms even in private.

(2) *Attitude- or belief-based persuasion*.[22] Hearing pro-trans statements and choices could affect participants' personal attitudes or beliefs about transgender workers. For example, a participant could become more willing to interact with a transgender worker after a group member tells an anecdote of a positive interaction with a transgender person.

(3) *Self-persuasion*. A participant may change her own attitudes towards transgender workers during the discussion, e.g., because she selectively searches for arguments in favor of her original view, and fails to fully adjust for this after the fact (**??**).

Below, I show experimental evidence in favor of inter-personal persuasion (consistent with (1) or (2)). I also present correlational evidence that favors norm-based persuasion, while indicating that attitude- or belief-based persuasion are unlikely to be key drivers.

5.2 Evidence consistent with persuasion

I first present evidence that is compatible with both forms of inter-personal persuasion (norm- and attitude-based), but is not supportive of self-persuasion. Additional suggestive evidence is presented in Appendix **??**.

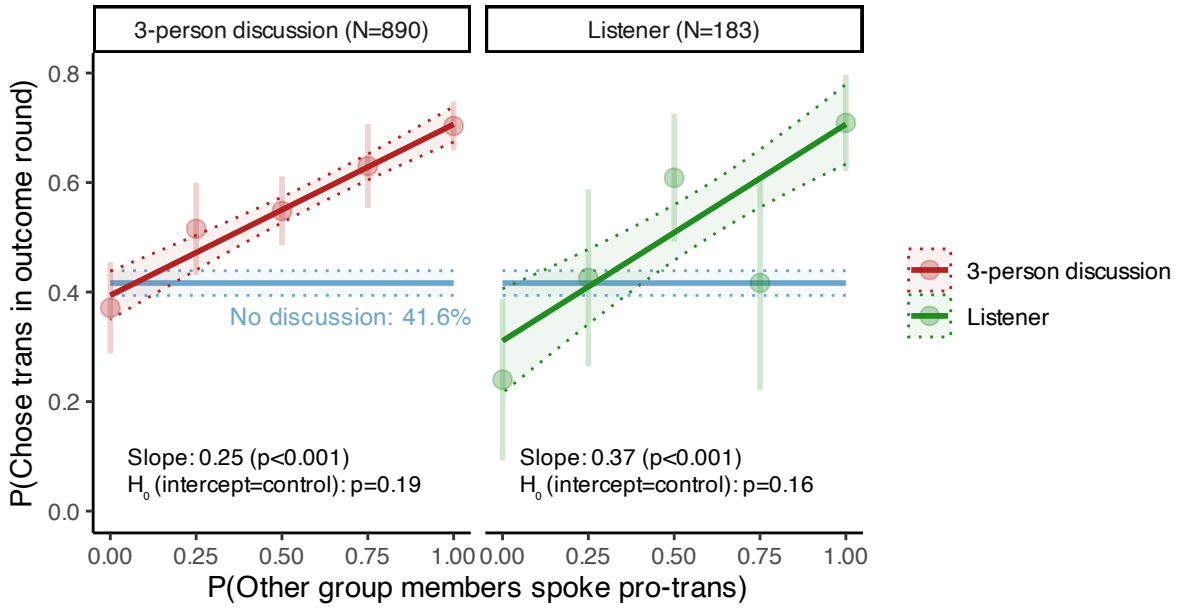
Effect of listening to discussion. Silent listeners in the *2-person discussion* treatment reduce discrimination as much as active discussants, suggesting that the effects are driven by what participants hear in the discussion from others (inter-personal persuasion), rather than through self-persuasion. Despite not speaking, listeners discriminate 13.3 p.p. less after the discussion ($p<0.001$, **??**), an effect statistically indistinguishable from both 2-person speakers ($p=0.86$) and 3-person discussion participants ($p=0.21$). The effects persist in the 2-9 week follow-up (β=11 p.p., $p<0.001$, **??**), and are similar for a short-run outcome that is completely private, i.e., unobserved by neighbors and enumerators (β=11 p.p., $p=0.01$, **??**, see Appendix **??**). These results suggest that the effects are maintained even when post-discussion behavior is private, consistent with persuasion mechanisms.

Heterogeneity by pro-trans discussion. Participants who hear more pro-transgender state-

[22] I denote both norm- and attitude-based influence as persuasion, since both mechanisms make very similar behavioral predictions: they both involve participants saying and choosing things that influence others to discriminate less *in private* after the discussion ends. This follows previous literature that defines persuasion as when a sender sends a message to a receiver with a potential interest in affecting the receiver's behavior (e.g., **?**), and work that includes *social proof* as a key mechanism for persuasion (**?**).

ments from others discriminate less afterward: each time a participant hears a group member speak positively about a transgender option (out of a maximum of 4, according to enumerator observations) is associated with a 6.2 p.p. reduction in discrimination after the discussion ($p<0.001$, **??**).[23] This pattern also holds for silent listeners (9.3 p.p., $p<0.001$, right panel), whose results are more likely to reflect causal effects since they have minimal opportunity to endogenously affect their group members' behavior. These correlational results are consistent with either norm-based persuasion (since more pro-trans discussions more strongly signal a pro-trans norm) or attitude or belief-based persuasion (since more pro-trans arguments are more likely to change attitudes or beliefs).

Figure 5: *More pro-trans discussions have more positive effects on post-discussion discrimination*



Notes: *P(Other group members spoke trans)* (x-axis) is the proportion of (discussion choice × participants) that spoke positively about a transgender option. There were two other group members, each with two opportunities to speak positively about a transgender worker. *P(Chose trans in outcome round)* (y-axis) is the probability a participant chose the transgender worker in the private outcome round. Plotted line of best fit and 95% confidence intervals are based on raw data with no controls. Left panel uses 3-person discussion participants in phases 1 and 2; right panel uses listeners in phase 2. Slope and p-values come from models with LASSO controls and all other controls in **??**. β=0.28 implies that each additional positive discussion from a group member corresponds to being 28/4 =6.2 p.p. more likely to select a transgender worker in the outcome round. H_0(*intercept=control*) tests whether a discussion participant for whom *P(Other group members spoke trans) = 0* exhibits different post-discussion discrimination from the control group.

5.3 Evidence for norm-based persuasion

Having shown that the results are driven by participants influencing *each other's* post-discussion behavior, I now document descriptive evidence in favor of norm-updating as the primary channel.

Effect of discussion on within-group norms. Discussions had large effects on participants' beliefs about their peers' behavior, providing strong evidence for norm-updating. In an

[23]The coefficients on pro-trans discussions are not significantly different when also controlling for the probability of hearing anti-trans discussions.

Table 4: *Effect of mechanism treatments (listener, public, observer)*

	Dep var: Chose trans (=1) (pairs with trans only)	
	Treatment round	Outcome round
	(1)	(2)
No discussion (public), pooled	−0.004	
	(0.021) [1.000]	
Observer (No discussion, public)		0.043***
		(0.026) [<0.001]
Non-observer (No discussion, public)		0.022***
		(0.031) [<0.001]
Speaker (2-person discussion)	0.170***	0.131***
	(0.028) [<0.001]	(0.028) [<0.001]
Listener (2-person discussion)		0.125***
		(0.033) [<0.001]
3-person discussion	0.210***	0.176***
	(0.033) [<0.001]	(0.031) [<0.001]
Num. observations	4432	4436
Num. participants	2216	2218
Num. groups	740	741
Outcome mean: no discuss (priv.), worker trans	0.49	0.41
p(Public=Speaker)	0.000	
p(Public=3-person discussion)	0.000	
p(Speaker=3-person discussion)	0.000	0.000
p(Observer=Non-observer)		0.000
p(Observer=Speaker)		0.000
p(Observer=Listener)		0.000
p(Observer=3-person discussion)		0.000
p(Speaker=Listener)		1.000
p(Listener=3-person discussion)		0.000

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes all treatment arms in phase 2 of data collection. In column (1), the *Listeners* are excluded (since they only make choices after they have observed the discussion). The specification used is otherwise the same as Tables **??** and **??**, columns (3). *No discussion (public), pooled* includes both *Observers* and *Non-observers*, since both these types are treated the same until after the treatment round.

incentivized prediction task, participants predicted whether their two group members selected transgender workers in private post-discussion choices (see Appendix **??**). The discussion increased these predicted probabilities by 24 p.p. (**??**, panel B, column 2, $p<0.001$). This change in perceived within-group norms appears to be a key mediator of the main treatment effects: controlling for within-group norms significantly weakens the effect on private discrimination (**??**), while controlling for other mechanism outcomes (community-wide norms, disapproval of discrimination, beliefs about reliability) does not. Section **??** will discuss explanations for why norms shift so starkly away from discrimination.

Effect of observing others' choices. To isolate whether norm-updating alone is sufficient to reduce discrimination, I examine the observers in the *No discussion (public)* treatment. Simply observing others' public choices (without a discussion) increases predictions that group members will select transgender workers in private by 4.5 p.p. (p=0.08, **??**, column 1), correcting participants' underestimation of discrimination (p-value of misperception = 0 is 0.17).[24] This corrected norm translates one-for-one into reduced discrimination in the private round (β=4.5 p.p., p=0.04, column 3).[25] In other words, simply observing others' choices without a discussion leads to small updates in perceived within-group norms, and this is sufficient to reduce later private discrimination. While this demonstrates that norm-updating can reduce discrimination, the discussions' effects are larger. This plausibly reflects the sources of asymmetry documented in Section **??**: for example, pro-trans participants' greater willingness to speak up likely leads to larger shifts towards an anti-discriminatory norm.

5.4 Suggestive evidence against attitude- or belief-based persuasion

The available descriptive evidence does not favor attitude- or belief-based persuasion.

Attitudes. Discussions produce only small or null effects on measures of personal attitudes towards transgender workers, indicating that effects are more likely to be operating through norm-updating than through changes in attitudes. The proportion of participants who said that (in their opinion) discrimination was wrong in response to two vignettes increased only slightly from 92.8% to 94.4% (p of difference: 0.05, **??**, panel B, column 2). Controlling for this measure of private attitudes does not weaken the effects of discussion on discrimination (p=0.92, **??**). There are also no effects on a double list experiment (**??**) that measured how many people agreed with the statement "In general, if I see a transgender person, I walk away" (p=0.22, **??**, panel B, column 1).

Beliefs about reliability. Belief-based (statistical) discrimination appears to underlie some of participants' unwillingness to select transgender people, driven by negative stereotypes that portray transgender workers as unreliable. However, the effect of the discussion does not appear to be driven by changes in such statistical discrimination. The discussion does not significantly affect beliefs about the reliability of transgender workers (p=0.15, **??**, panel

[24] There are no average effects of making choices public *in the treatment round* (β=-0.4 p.p., p=0.83), so observing public choices in the *No discussion (public)* arm is on average equivalent to observing control group's choices.

[25] Using the cross-sectional heterogeneity in whether participants observed pro- or anti-trans choices further shows that observing one additional group member's choice to select a transgender worker (out of a maximum of 4) increases the probability of a selecting transgender worker in the private outcome round by 26 p.p. ($p<0.001$, column 4).

Table 5: *Effect of observing others' choices*

	Predicted prob. of others selecting trans (within-group, outcome round)		Chose trans in private outcome round (=1)	
	(1)	(2)	(3)	(4)
Observer (No discussion, public)	0.045*	−0.176***	0.045**	−0.091**
	(0.026)	(0.042)	(0.022)	(0.039)
	[0.079]	[<0.001]	[0.038]	[0.022]
π_{-i} = P(others in group selected trans in treatment round)		0.034		−0.032
		(0.038)		(0.037)
		[0.374]		[0.385]
π_{-i} × Observer (No discussion, public)		0.434***		0.258***
		(0.077)		(0.069)
		[<0.001]		[<0.001]
Num. observations	3487	3487	3516	3516
Num. participants	1748	1748	1758	1758
Num. groups	654	654	654	654
Outcome mean: No discuss. (priv.)	0.36	0.36	0.42	0.42
Controls	X	X	X	X

Notes: Standard p-values are in brackets. Unit of observation is the participant × choice level. Outcome in columns 1–2 is the predicted probability that other group members select a transgender worker in private. Outcome in columns 3–4 is whether participant selected a transgender worker themselves in the outcome round. In all columns the sample is only the *Observers (no discussion, public)* and the control group (*No discussion, private*) in phase 2. *P(others in group selected trans in treatment round)* (π_{-i}) is the proportion of times (out of a maximum of 4) that the other two participants in the group selected a transgender worker in the treatment round. All regressions control for *P(selected trans in treatment round*, the proportion of times (out of a maximum of 2) that the participant herself selected a transgender worker in the treatment round. Other controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; relative # items offered by the transgender worker; relative reliability score; and a dummy for whether the reliability score was shown.

A, column 3); controlling for these beliefs does not weaken the effect on discrimination (**??**); and I find no evidence that the 3-person discussion reduces the belief-based component of discrimination, although I am underpowered for this latter test (**??**, column 2).

6 Mechanisms: sources of asymmetry

Having shown evidence in favor of norms-based persuasion as a key channel for the effects on post-discussion behavior, I now examine why discussions produce a large average shift away from discrimination.

Standard models of communication in economics predict that communication should lead behavior to converge, and (absent misperceptions in participants' priors) should not change average behavior.[26] The substantial reduction in discrimination therefore requires asymmetries in the discussion process. I identify three key asymmetries: pro-transgender participants are disproportionately vocal, participants fail to correctly account for this selective communication when inferring norms, and moral arguments against discrimination appear to be more persuasive than anti-trans arguments. Much of the evidence on these asymmetries comes from

[26]Suppose each participant i is learning about their own (dis)utility θ_i from hiring a transgender worker. Under a standard Bayesian communication model (with a common prior and truthful information exchange), participants converge to a common posterior (**??**). The martingale property of Bayesian updating also implies that the posterior belief (and therefore post-communication behavior) will not shift on average across participants relative to the prior *if* the prior is correctly specified, i.e., it accurately represents the true data-generating process.

heterogeneity analyses rather than exogenous mechanism treatments; these results should be interpreted with some caution.

6.1 Behavior during discussion and willingness to speak up

The first asymmetry I document is that pro-transgender participants are more vocal than others, leading to discussions that were strongly in favor of transgender workers. Their stronger advocacy appears to be partly driven by a stronger desire to influence other group members' choices.

Choices during the discussion (treatment round). The choices made *during* the discussion are favorable towards transgender workers. Participants were 20 p.p. more likely to select a transgender worker in the collective choices during the discussion than in the private choices made by *No discussion (private)* participants (**??**, $p<0.001$). In the discussion choices, there was even positive discrimination in favor of transgender workers relative to non-transgender workers (11 p.p., $p<0.001$).

Pro- and anti-trans statements in discussion. Participants also on average communicated about transgender workers in a positive way. Based on transcripts with a sentence-level indicator of arguments in favor of or against the transgender worker, participants were 2.5x more likely to make pro-trans than anti-trans statements (38% of sentences were pro-trans, vs. 15% of sentences were anti-trans, p of difference <0.001).

Substantial minority of pro-trans participants. A substantial minority of participants favor transgender workers despite strong average discrimination. **??** presents individual-level estimates of willingness to pay (WTP) for transgender workers in the outcome round, derived from a regression with individual fixed-effects. The distribution reveals wide variance, so despite a substantial negative WTP on average, 31% of control participants have positive WTP for transgender workers. This sizable pro-transgender faction is likely to be important for explaining the pro-transgender nature of group discussions.

Pro-trans participants are more vocal. This pro-trans pattern in the discussion appears to be partly explained by pro-trans individuals choosing to be more vocal. I use private choices after the discussion as a proxy for baseline pro-trans behavior and show that this is correlated with dominating the discussion of transgender workers (**??**).[27] For each additional transgender worker a participant selected in the private round, they were 11 p.p. more likely to speak first (32%, p=0.03) and 15 p.p. more likely to dominate discussions (27%, p=0.02) when discussing transgender workers (according to enumerator observations). No such correlation exists for non-transgender worker discussions, implying that pro-trans participants selectively spoke up about transgender workers, despite not being more talkative in general. This asymmetry could plausibly be a driver of reductions in post-discussion discrimination.

[27] Since I purposely did not collect baseline measures of discrimination (to minimize priming and experimenter demand effects), the evidence here should be taken as suggestive. However, under a reasonable monotonicity assumption, the post-discussion discrimination will correlate positively with baseline discrimination: as long as participants are not too effective at persuading each other, initially anti-trans participants are unlikely to become more pro-trans than those who started off as pro-trans from the beginning.

6.1.1 *Why do pro-trans participants speak up more?*

Pro-trans participants might have spoken up more because of *social image concerns* (e.g., anti-trans participants self-censor because they didn't want to appear discriminatory), or because they had greater *influence concerns* (e.g., pro-trans participants cared more about stopping others taking discriminatory actions) (see **?** for a framework describing these motives). Suggestive evidence from mechanism outcomes indicate that influence concerns were likely more important in driving pro-trans participants' advocacy.

Social image concerns. There is no strong evidence for social image concerns driving pro-trans participants' disproportionate willingness to speak up. First, the *No discussion (public)* treatment did not reduce discrimination in the treatment round (**??**, column 1, p=0.83), implying that exogenously increasing social image concerns by revealing choices to other group members did not reduce discrimination.[28] This suggests that social image concerns are unlikely to be a key driver of pro-trans communication in the discussions, although importantly I cannot rule out that social image concerns have a stronger effect in discussions than when silently making choices (e.g., because the social costs of speaking are larger than choosing, or because of dynamically increasing social image concerns after pro-trans statements). Second, pro-trans advocates were not more likely to report making statements "that others would agree with" or in order "to look good" in post-discussion debriefs compared to those who didn't advocate for transgender workers (**??**, bottom panels, p= 0.76 and 0.35).

Influence concerns. The evidence instead hints at an aim to persuade others to change their behavior. Participants who spoke in favor of transgender workers were 8.4 p.p. more likely to report trying to get others to agree with them (13%, p=0.03, **??**, top left). While these differences are only correlational and do not extend to all measures of persuasion (**??**, top right), they suggest pro-trans participants disproportionately engaged in the discussion to shape others' choices. Pro-trans participants' behavior may be more driven by persuasion concerns than anti-trans participants for a number of reasons: they may have a more intense preference in favor of transgender workers, or they may care more about others' behavior (e.g., if they perceive discrimination as a social wrong, rather than a personal preference). Alternatively, pro-trans participants may recognize their rhetorical advantage (**?**): as I show below (Section **??**), they deploy moral arguments that may be especially compelling, thus making their active participation in the discussion more attractive.

6.2 *Incorrect inference about the norm*

The second source of asymmetry that could explain the large shift towards an anti-discriminatory norm is that people did not make correct inferences, tending to *over-update* on pro-trans behavior and to *under-update* on anti-trans behavior during the discussion. For example, if participants did not account for pro-trans speakers' disproportionate willingness to speak up, the discussion would make them overestimate their group's true support for transgender workers. This mechanism builds on theories of misperceived norms (**?**). But here, the norm is

[28] The null effect is not driven by participant inattentiveness, because participants' choices in *No discussion (public)* did react: they converged within groups, despite no mean effects (p-value of effect on intracluster correlation within groups: 0.10).

misperceived because of the conversation itself. I find two main pieces of suggestive evidence of such incorrect inference.

Silence neglect. Participants appear to have neglected the informational value of silence during discussions, accentuating updates toward pro-transgender norms. While participants strongly updated their beliefs after hearing pro-transgender statements, they failed to interpret silence as evidence of anti-transgender attitudes. When no group members made pro-transgender statements, participants' predictions about their peers' choices remained indistinguishable from control group predictions (p of difference: 0.22, **??**). This implies that participants did not interpret silence as a signal of anti-trans attitudes, even though silence often indicated opposition rather than indifference to selecting transgender workers. The failure to interpret silence as informative carried through to post-discussion behavior: 3-person discussion participants who heard no pro-trans statements did not increase discrimination relative to the control group (**??**).[29] This evidence is in line with theories of the "spiral of silence" in which people come to believe that a view is more prevalent than it truly is because of selective communication (**??**), and work on selection neglect showing that people fail to discount how information is filtered before reaching them (**?**).[30]

Overcorrection of misperception. The evidence is consistent with the discussion's effects partly stemming from correcting—and then overcorrecting—initially misperceived norms. I do find evidence of an initial misperception: the control group underestimated how often their peers would privately choose transgender workers (by 5.3 p.p., $p<0.001$, **??**). But discussions did more than fix this misperception—they created a new one in the opposite direction. After discussions, participants overestimated their peers' selection of transgender workers by 3.0 p.p. (p=0.04). This pattern is also consistent with a "spiral of silence" mechanism: pro-trans participants speak up more, other participants do not sufficiently account for this, and thus those other participants overestimate support for transgender workers among their peers.

6.3 Rhetorical asymmetry and moral arguments

The discussion transcripts reveal a third source of asymmetry: the types of arguments participants used, and how persuasive they appear to be. Pro-trans advocates more frequently invoked moral concepts (like equality, rights, and opportunity) when discussing transgender workers, and this moral framing is strongly associated with lower post-discussion discrimination. While this pattern is only correlational, it suggests a form of *rhetorical asymmetry* (**?**) in which pro-transgender arguments were more persuasive than anti-transgender ones. The pattern emerges consistently across multiple analyses: a sentence-level transcript analysis, AI-generated hypotheses about discussion characteristics, and enumerator observations of discussions. The relationship holds for silent listeners, for which causality is more plausible, and even after controlling for within-discussion choices. These findings align with psycho-

[29] For listeners, the raw data suggest that discussions with no pro-trans statements may have slightly decreased the predicted and actual probability of choosing a trans worker (**??** and **??**, right panels). The intercept is not significantly different from 0 in either case when controls are included (p=0.16 and 0.22), but the coefficient is negative. While the sample size of listeners is small, this may indicate *partial* rather than total silence neglect.

[30] Unlike the standard spiral-of-silence mechanism (that emphasizes social image concerns), participants' desire to persuade others appears to be the primary driver of selective communication in this context.

logical research on moral reframing, which shows that arguments framed in moral terms are particularly persuasive when they align with listeners' values (**??**). Combined with the evidence on norm-updating, this suggests that moral arguments may serve as powerful signals of anti-discriminatory norms.

Transcript sentences. Sentence-level semantic analysis of the 3-person discussions suggests that moral language predicts reductions in post-discussion discrimination. Using OpenAI's *embedding-3-small* model, I use *k*-means clustering to classify sentences based on similarity into 20 groups (**??**, see Appendix **??** for methodology). Sentences about "giving opportunities" to a worker and sentences that explicitly acknowledge that a worker is transgender are especially common for choices involving a transgender worker.[31] These sentences also strongly predict selecting transgender workers privately after the discussion: a 1 SD increase in "giving opportunities" sentences corresponds to a 6.8 p.p. decrease in post-discussion discrimination (q=0.001, cluster 1), and a 1 SD increase in explicitly acknowledging transgender identity predicts a 5.7 p.p. increase (q=0.001, cluster 2). In contrast, discussions that focused on workers' appearance predict weaker treatment effects (β=-7.3 p.p., q=0.003, cluster 3), in line with stereotypes portraying transgender people as looking strange or untrustworthy.

Broader transcript features. To analyze higher-level characteristics of the discussion that are correlated with reduced discrimination, I follow recent work that uses AI to generate hypotheses (**??**). Specifically, I (i) use a large language model (LLM) to generate 500 hypotheses about the main difference between two randomly selected transcripts (without revealing any post-discussion outcomes); (ii) use an LLM to rate all transcripts based on how much they satisfy each hypothesis; (iii) combine hypotheses whose ratings are highly correlated using factor analysis; and (iv) test whether the aggregate rating is correlated with post-discussion discrimination (see Appendix **??**).

Using this method, I find that discussions where the transgender worker's identity is explicitly acknowledged and with a stronger focus on social equity and support for marginalized individuals are strongly associated with less discrimination (q<0.001, **??**, rows 1 and 2). A focus on gender identity and on grocery items during the transgender discussion is also associated with less post-discussion discrimination (q<0.05, rows 3 and 4). In contrast, discussions about transgender workers that focus on other details about the worker (e.g., the language proficiency or delivery performance) are associated with weaker treatment effects (q=0.07 and 0.03), in line with the view that moral arguments are more persuasive than practical ones in this context. Notably, there do not seem to be moral *anti*-transgender arguments that are persuasive: anti-trans participants only resort to details about the workers' qualifications.

Enumerator observations. Enumerators' observations are consistent with this picture. When discussing transgender workers, participants more frequently cite moral rationales (e.g., wanting to give an opportunity to or help the worker; saying that the worker is "also human") compared to discussions of non-transgender workers (difference = 32.5 p.p., p<0.001, **??**). And

[31] Indeed, the transcript quotes that were *most* uniquely representative of transgender discussions (relative to non-transgender discussions, based on relative cosine similarity of embeddings) were "*They are transgender, therefore, they deserve an opportunity*" and "*We should give them equal rights too*". Participants were especially likely to use words like "opportunity", "chance", and "rights" (see word cloud in **??**).

these moral rationales correlate with lower discrimination after the discussion: listeners who heard them were 6 p.p. more likely to select transgender workers afterward ($p<0.001$).[32]

Moral discussions. The evidence above suggests that discussions focusing on giving opportunities, social equity, or other moral notions are associated with reductions in post-discussion discrimination, in keeping with the psychological literature on moral reframing (**??**). To more directly test the hypothesis that moral reframing reduces discrimination, I use an LLM to rate how much each discussion transcript invoked morality on a 0–1 scale (see Appendix **??**). Participants used moral language substantially more when discussing transgender workers (from 0.18 to 0.44, +0.99 SD, $p<0.001$). And this moral language is highly predictive of post-discussion discrimination: a 1 SD increase in morality rating corresponds to being 39 p.p. more likely to select transgender workers after the discussion ($p<0.001$, **??**, column 1), even when controlling for general pro-trans statements (column 2) and choices (column 3). While this evidence is only correlational, it suggests that the moral content of discussions may reduce discrimination independently of the within-discussion choices. Moral language moreover appears to operate through the *norm-updating* channel described above, i.e., by affecting participants' perceptions of the norm of behavior towards transgender workers: a 1 SD more moral discussion is also associated with large updates in within-group norms (14 p.p. when controlling for treatment round choices, $p=0.02$).

Heterogeneity on persuasiveness. One piece of suggestive evidence against persuasion and rhetorical asymmetry is that participants who rate their group members as more persuasive in a pre-discussion question do not discriminate less afterward. Neither listeners nor discussion participants discriminate less when their peers score higher on a pre-discussion persuasiveness index ($p=0.61$ and 0.91, **??**). This measure has limitations—it captures perceptions of a persuasive personality before the discussion, rather than transgender-specific advocacy or likelihood of speaking up. The null result thus doesn't preclude effects stemming from pro-trans participants being more willing to speak up or from selecting more persuasive moral arguments specifically when discussing transgender workers. Nevertheless, the null result suggests caution in attributing effects entirely to rhetorical asymmetry.

7 Discussion

A group discussion about hiring transgender workers sharply reduced discrimination in subsequent real-stakes hiring choices. Despite lasting only 10 minutes, the discussion also decreased discriminatory choices several weeks later. I show experimental evidence that the impact is driven by inter-personal persuasion, and suggestive evidence based on correlations that it stems from pro-transgender participants speaking up more frequently during discussions, leading others to perceive social norms as less discriminatory and consequently reduce their own private discrimination after the discussion ended. These results are a proof-of-concept that the horizontal communication can lead to large reductions in discrimination. Progressive social change can thus occur not only by generating social contact between in-groups and out-groups,

[32]The increase in pro-social reasoning translates to participants' reported reasoning in the private outcome round. When participants were asked why they made their outcome-round choices, those who had been involved in a discussion were 11 p.p. (24%) more likely than the control group to cite pro-social reasons for their choices ($p=0.01$).

Table 6: *Morality rating of discussion predicts reduced post-discussion discrimination*

	Dep var: Chose trans in outcome round (=1)		
	(1)	(2)	(3)
3-person discussion	−0.008	−0.094***	0.027
	(3.3×10^{-2}) [0.799]	(3.5×10^{-2}) [0.008]	(2.7×10^{-2}) [0.327]
Discussion morality score, trans choices (Z)	0.394***	0.287***	0.116**
	(5.4×10^{-2}) [<0.001]	(5.8×10^{-2}) [<0.001]	(4.8×10^{-2}) [0.016]
Discussion morality score, non-trans choices (Z)	−0.016	−0.007	0.024
	(7.3×10^{-2}) [0.831]	(7.2×10^{-2}) [0.924]	(6.8×10^{-2}) [0.728]
P(others in group spoke pro-trans in discussion)		0.215***	
		(4.3×10^{-2}) [<0.001]	
Num. times group chose trans in treatment round			0.214***
			(1.1×10^{-2}) [<0.001]
Num. observations	4362	4362	4362
Num. participants	2249	2249	2249
Num. groups	751	751	751
Outcome mean: No discuss. (priv.)	0.42	0.42	0.42

Notes: Sample includes *3-person discussion* and *No discussion (private)* participants, so the control group is the omitted category. *Discussion morality score* is the morality score of the discussion transcripts rated by an LLM, designed to measure how much morality is invoked (see Section **??** and Appendix **??**). Each discussion-choice transcript receives a rating. I average these ratings for each group's transgender and non-transgender choices separately, normalizing to ensure mean of 0 and standard deviation of 1. *Others in group: proportion of discussions spoke in favor of trans* is the proportion of times (out of a maximum of 4) the other group members spoke positively about a transgender option. *Num. times group chose trans in treatment round* is the number of times transgender worker was chosen by the group in the treatment round. All 4 of these additional predictors are coded as 0 for the control group.

but also by encouraging communication among the in-group in the right circumstances.

Why did horizontal communication sharply reduce discrimination in this setting? It can also amplify prejudice, as in 1930s Germany when anti-Semitic sentiment spread through social networks (**?**), or in other psychological studies examining group dynamics (**?**). The norm-based persuasion mechanism I show evidence for is directionally neutral – it could either reduce or increase discrimination depending on which voices dominate the discussions. The central puzzle is thus explaining why pro-trans participants were so vocal in this context.

The contextual factor that appears to be key is a striking disconnect: while discrimination is strong and common in the absence of communication, there is nevertheless widespread agreement that discrimination is wrong (e.g., 93% of control participants say it is wrong). This contrasts with many settings where discrimination is not only common but also widely accepted or even morally encouraged. This gap between behavior (descriptive norms) and moral beliefs (prescriptive norms) could explain four key patterns that could drive the positive effects: (i) the substantial minority of pro-trans participants; (ii) the minimal social costs for advocating against discrimination, unlike in contexts where discrimination is both practiced and accepted; (iii) pro-trans participants' greater propensity to try to influence others, consistent with them viewing discrimination as a moral wrong rather than a personal choice, and thus caring more about others' actions; and (iv) the particular prevalence and influence of pro-trans moral arguments about equal rights and fairness that align with participants' stated values (unlike in Nazi Germany, where moral narratives focused on impurity were used to *encourage* discrimination). All of these factors would, in turn, contribute to pro-trans participants'

disproportionate willingness to speak up, and thus plausibly explain the reductions in post-discussion discrimination. They could also explain why the descriptive norm seems to be so responsive to the discussion. Future research could valuably explore the boundary conditions of when horizontal communication is helpful in other contexts. Looking for other contexts with a wedge between the level of discrimination (what people do) and the acceptance of discrimination (what people say is right), for example, using attitude surveys, would be a promising avenue for finding such contexts.

The study has several important limitations. First, the main effects are only measured shortly after the discussions, which may explain the large coefficients, and the medium-run effects are only based on hypothetical choices. The post-discussion decisions are also of the same type as those made during the discussions. This raises the concern that some of the effects may be driven by participants' drive to maintain consistency (although the large effects on silent listeners alleviate this concern, since they did not make choices in the discussion). Second, I do not elicit baseline measures of discrimination. While this was done deliberately to minimize the risk that participants guessed the purpose of the experiment and reduce experimenter demand effects, it limits the ability to conclusively understand which types of participants speak up and how group composition relates to persuasion. Third, the communication involved in the discussion is not entirely natural: the presence of an enumerator facilitating the discussion may influence participants' behavior, and participants make incentivized choices at the same time as communicating. Although facilitator influence may also exist for the control group, and the intervention resembles a policy-style intervention, the results may not as clearly map onto the effects of organic horizontal communication in society absent an intervention. Fourth, while the evidence in favor of inter-personal persuasion is based on an experimental variation, the evidence on other mechanisms related to norm updating and asymmetries is only suggestive and primarily based on consistent patterns of correlations. Future research could explore participants' decisions to speak up in more organic scenarios (e.g., with no facilitator), and use exogenous variation in group composition and baseline attitude measures to better understand mechanisms.

Future research could also explore how to design policies based on the insight that in-group communication can reduce discrimination under the right circumstances. Policies that create discussions at scale to change attitudes are particularly promising (e.g., see **???**). The role of vocal pro-trans participants also suggests it may be possible to amplify the voice of pro-social actors within a network by identifying them and training them to influence others' behavior. This tactic underlies the tactic followed by activist NGOs, and has also shown promise when trying to reduce the stigma surrounding menstruation (**?**). Finally, we could use groups rather than individuals to make high-stakes decisions in order to reduce discrimination in cases where a pro-minority prescriptive norm is in place.

A Additional tables and figures

Figure A1: *Experimental design (detailed)*

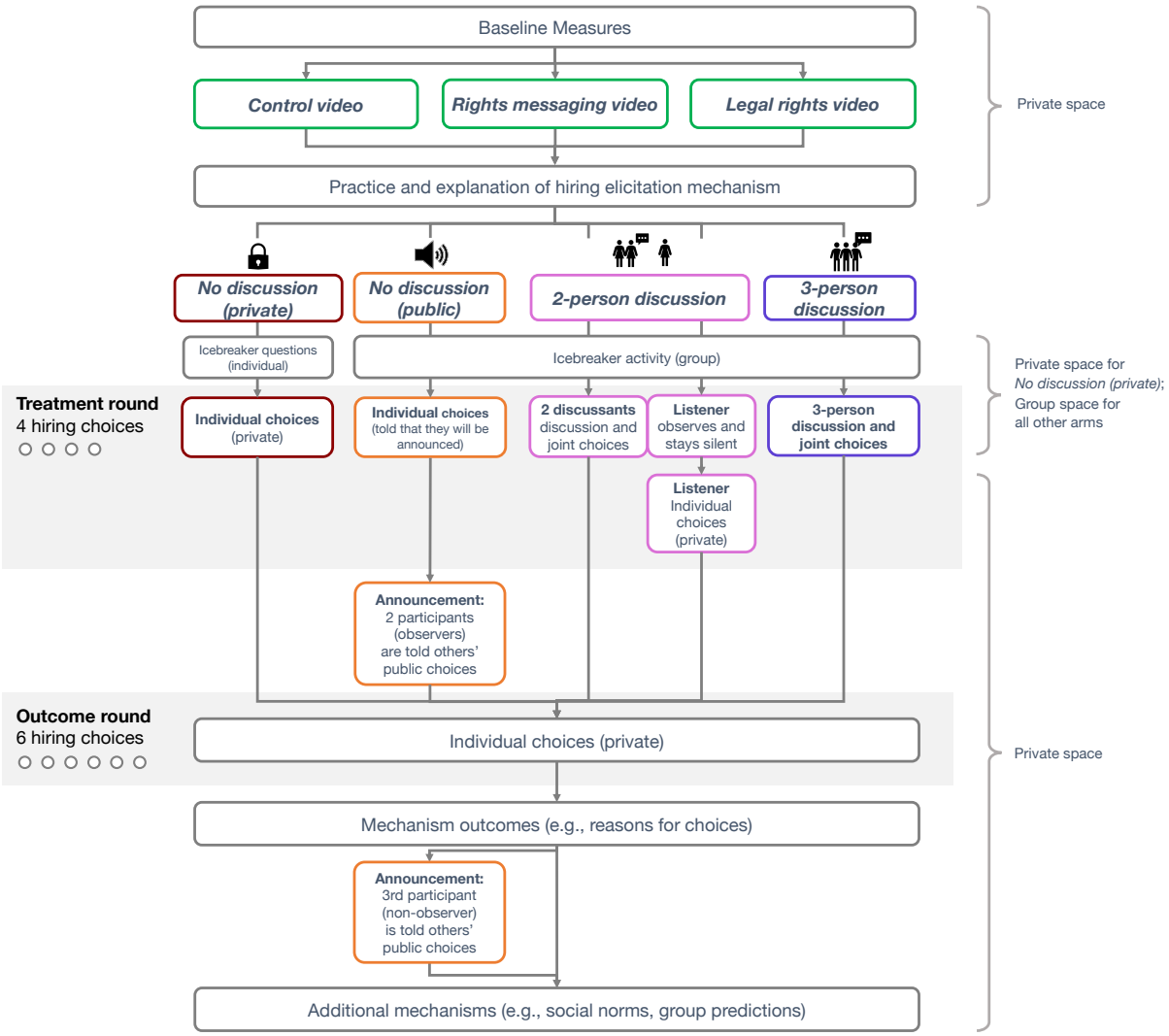
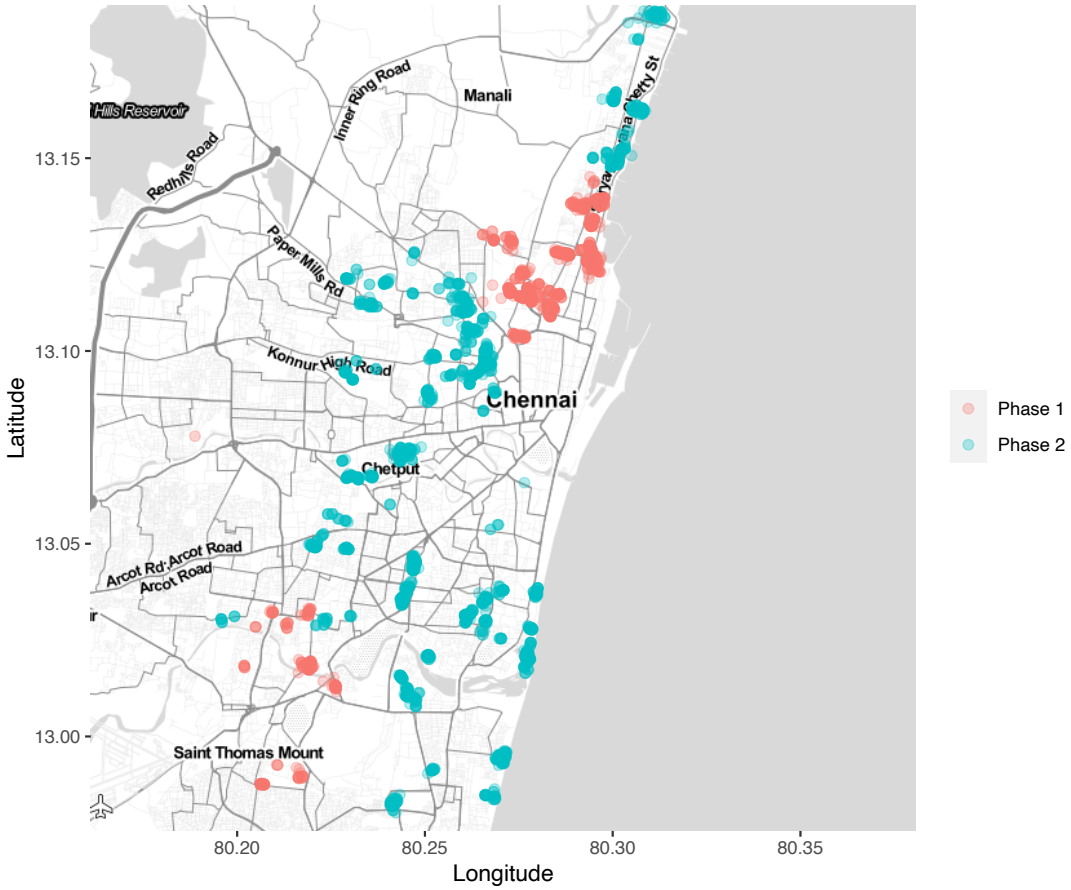


Figure A2: *Survey locations*



Notes: This shows the location of each survey. Red dots denote surveys from phase 1. Blue dots denote surveys from phase 2.

Table A3: *Transgender photo recognition confusion matrix*

	Correct gender:		
Participant guess	Male or female	Transgender	Total
Male or female	1239	10	1249
Transgender	15	332	347
Total	1254	342	1596

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. From supplementary data collection that took place in August-September 2022 (N=114). Each participant was shown 14 worker photos. 11 of these were male or female, and 3 were transgender. The participant was asked to select all the photos that were transgender. Transgender photos were recognized as being transgender 97% of the time (332/342), and non-transgender photos were falsely identified as transgender photos only 1.2% of the time (15/1254).

Table A4: *Balance for 3-person discussion (Phases 1 + 2)*

	Means		p-values	
	(1)	(2)		
Variable	No discussion (private)	3-person discussion	p-value (1)=(2)	
Age	35.60	35.80	0.68	
Female (=1)	0.85	0.86	0.40	
Speaks English (=1)	0.14	0.14	0.96	
Reads English (=1)	0.26	0.25	0.47	
Hindu (=1)	0.84	0.84	0.81	
Bachelor's degree (=1)	0.20	0.18	0.37	
Married (=1)	0.84	0.83	0.82	
Employed (=1)	0.22	0.22	0.61	
Landlord (=1)	0.09	0.07	0.18	
Num. children	0.64	0.64	0.88	
Employer (=1)	0.25	0.21	0.01	***
Household size	4.19	4.16	0.54	
Monthly h.h. food expenditure per capita (Rs.)	2310.04	2309.80	1.00	
F-test: statistic			**0.89**	
F-test: p-value			**0.56**	

Notes: Columns 1 and 2 show the means of the covariates for the *No discussion (private)* arm and *3-person discussion* arm, including participants from phases 1 and 2. Column 3 shows the *p*-value of a test of the equality of columns 1 and 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The base of the table displays the test statistic and p-value for an F-test for the equality of all covariates across the treatment arms.

Table A5: *Balance for transgender rights videos*

	Means			p-values			
	(1)	(2)	(3)				
Variable	Control video	Rights messaging video	Legal rights video	p-value (2) - (1)		p-value (3) - (1)	
Age	34.50	34.38	35.36	0.81		0.11	
Female (=1)	0.86	0.85	0.85	0.83		0.79	
Speaks English (=1)	0.19	0.18	0.18	0.60		0.79	
Reads English (=1)	0.35	0.32	0.31	0.17		0.10	
Hindu (=1)	0.80	0.80	0.82	0.89		0.46	
Bachelor's degree (=1)	0.26	0.25	0.22	0.75		0.05	*
Married (=1)	0.82	0.84	0.85	0.28		0.21	
Employed (=1)	0.22	0.21	0.20	0.82		0.33	
Landlord (=1)	0.09	0.10	0.10	0.59		0.35	
Num. children	0.69	0.69	0.68	0.89		0.50	
Employer (=1)	0.33	0.32	0.32	0.59		0.81	
Household size	4.11	4.25	4.20	0.06	*	0.19	
Monthly h.h. food expenditure per capita (Rs.)	2373.21	2232.74	2278.81	0.05	*	0.18	
F-test: statistic				**0.54**		**0.92**	
F-test: p-value				**0.90**		**0.54**	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns 1-3 show the means of the covariates for each of the rights videos arms. Columns 4-5 show the *p*-value of a test of the equality of columns 1-3. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The base of the table displays the test statistic and p-value for an F-test for the equality of all covariates across the treatment arms.

Table A6: *Balance for phase 2 discussion arm treatments*

	Means				p-values		
	(1)	(2)	(3)	(4)			
Variable	No discussion (private)	No discussion (public)	2-person discussion	3-person discussion	p-value (1)=(2)	p-value (1)=(3)	p-value (1)=(4)
Age	35.13	34.02	35.12	34.57	0.04 **	1.00	0.44
Female (=1)	0.85	0.84	0.86	0.85	0.67	0.60	0.80
Speaks English (=1)	0.19	0.19	0.17	0.18	0.90	0.23	0.66
Reads English (=1)	0.32	0.36	0.30	0.32	0.16	0.39	0.81
Hindu (=1)	0.80	0.82	0.81	0.81	0.33	0.78	0.62
Bachelor's degree (=1)	0.24	0.26	0.22	0.27	0.23	0.57	0.28
Married (=1)	0.83	0.84	0.85	0.82	0.80	0.38	0.70
Employed (=1)	0.21	0.19	0.21	0.24	0.31	0.77	0.39
Landlord (=1)	0.09	0.09	0.10	0.09	0.98	0.63	0.94
Num. children	0.67	0.70	0.70	0.68	0.28	0.24	0.81
Employer (=1)	0.33	0.33	0.32	0.30	0.78	0.92	0.32
Household size	4.21	4.21	4.12	4.24	0.99	0.22	0.72
Monthly h.h. food expenditure per capita (Rs.)	2304.84	2249.57	2345.54	2264.07	0.45	0.60	0.65
F-test: statistic					**0.89**	**0.47**	**0.36**
F-test: p-value					**0.56**	**0.94**	**0.98**

Notes: Columns 1-4 show the means of the covariates for all discussion-treatment arms in Phase 2. Columns 5-7 show the *p*-value of a test of the equality of columns 1-4. * p < 0.1, ** p < 0.05, *** p < 0.01. The base of the table displays the test statistic and p-value for an F-test for the equality of all covariates across the treatment arms.

Table A7: *Discussion effects are robust to restricting to control video only*

	Chose worker in outcome round (=1)		Chose trans in outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans × 3-person discussion	0.205***	0.198***	
	(0.041) [<0.001]	(0.040) [<0.001]	
Worker is trans	−0.245***		
	(0.024) [<0.001]		
3-person discussion	−0.013	−0.001	0.188***
	(0.019) [0.486]	(0.017) [0.968]	(0.035) [<0.001]
Num. observations	4530	4530	1510
Num. participants	755	755	755
Num. groups	252	252	252
Outcome mean†: no discuss (priv.), worker non-trans	0.62	0.62	
Outcome mean: no discuss (priv.), worker trans	0.37	0.37	0.37
Controls		X	X
Controls interacted with worker is trans		X	

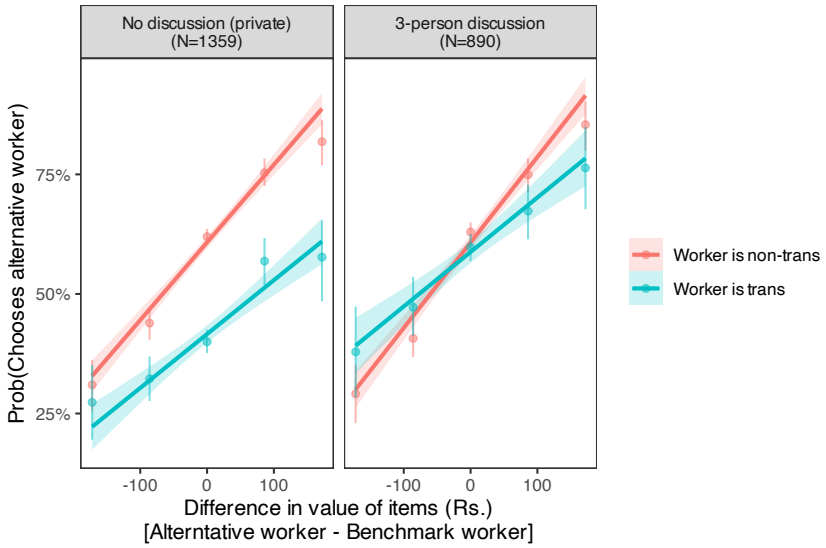
Notes: Sample includes *only* participants who saw the control video, and excludes participants who saw the rights messaging or legal rights videos. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. Column (3) only includes choices that involved a transgender worker. In columns (1) and (2), the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was chosen in the private choices in the *outcome round*. In column (3), it is whether the transgender worker was selected. *Worker is trans* $= 1$ when the alternative worker is transgender, and is 0 when the alternative worker is male or female. The specification used is seen in equation **??**. Controls include stratum fixed effects; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; and the controls selected by double LASSO (see Section **??**). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown. Columns (2) and (3) also include controls for the relative # items offered by the alternative worker, the relative reliability score of the worker, and a dummy for whether the reliability score was shown.

Table A8: *Medium-run effects of videos on hypothetical hiring choices (2-9 weeks)*

	Chose worker in follow-up round (=1)		Chose trans in follow-up round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.121***		
	(0.015) [<0.001]		
Rights messaging video	0.003	0.003	0.015***
	(0.011) [1.000]	(0.011) [1.000]	(0.020) [<0.001]
Legal rights video	−0.002***	−0.001	0.028***
	(0.011) [<0.001]	(0.010) [1.000]	(0.019) [<0.001]
Worker is trans × Rights messaging video	0.012	0.012	
	(0.021) [0.500]	(0.021) [0.500]	
Worker is trans × Legal rights video	0.029***	0.028***	
	(0.021) [<0.001]	(0.021) [<0.001]	
Num. observations	19 266	19 266	6416
Num. participants	3230	3230	3224
Num. groups	1134	1134	1133
Outcome mean: control vid., worker non-trans	0.62	0.62	
Outcome mean: control vid., worker trans	0.49	0.49	0.49
Controls		X	X
Controls interacted with worker is trans		X	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes all participants. Controls include dummies for the discussion-arm treatments, as well as the other controls specified in Table **??**. In the follow-up survey, workers in a pair always had the same reliability score and offered the same number of items. Specification is otherwise the same as Table **??**.

Figure A9: *Inferring WTP to avoid transgender workers*



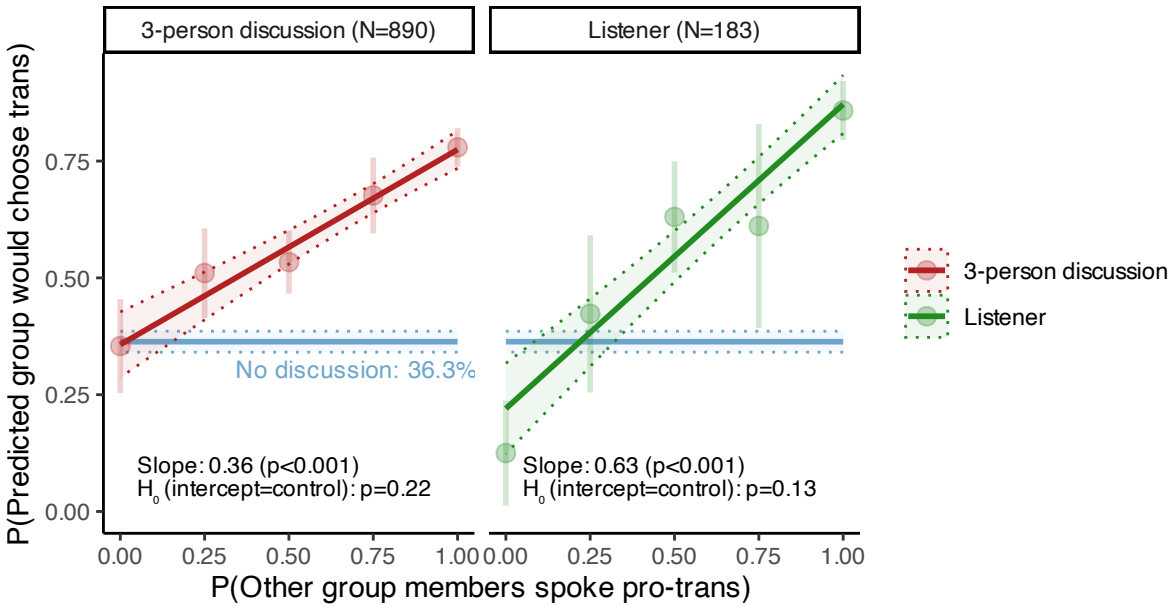
Notes: Points represent the probability of choosing the alternative worker at the given difference in value of items in Rs. Solid lines represent a linear fit. I take the reduction in probability that an option is chosen when a worker is transgender in each treatment group, and divide it by the gradient of selecting an option with respect to item value. Gradient with respect to item value (pooled across all treatment groups and alternative worker types) is 0.0015, implying that increasing the value of the items offered by an option A by 100 Rs. (relative to the other option B in the pair) increases the probability of a participant selecting A by 15 p.p. The mean reduction in the probability of choosing the alternative worker when they are trans is 0.19 in the control group, and 0.02 in the discussion group. This corresponds to a willingness to pay to not choose transgender workers of 0.19 / 0.0015 = 127 Rs. in the control group that reduces to 0.02 / 0.0015 = 13 Rs. in the discussion group.

Table A10: *Dominated and dominating choices: negative discrimination decreases and positive discrimination increases*

	Dep var: Chose trans worker (=1) (pairs with trans only)		
	Trans dominates	Trans is dominated	Neither dominates
	(1)	(2)	(3)
3-person discussion	0.175*** (0.031) [<0.001]	0.127*** (0.032) [<0.001]	0.179*** (0.026) [<0.001]
Num. observations	1088	1185	2225
Num. participants	960	1063	1721
Num. groups	600	628	735
Outcome mean: no discuss (priv.), worker trans	0.53	0.35	0.40
Controls	X	X	X

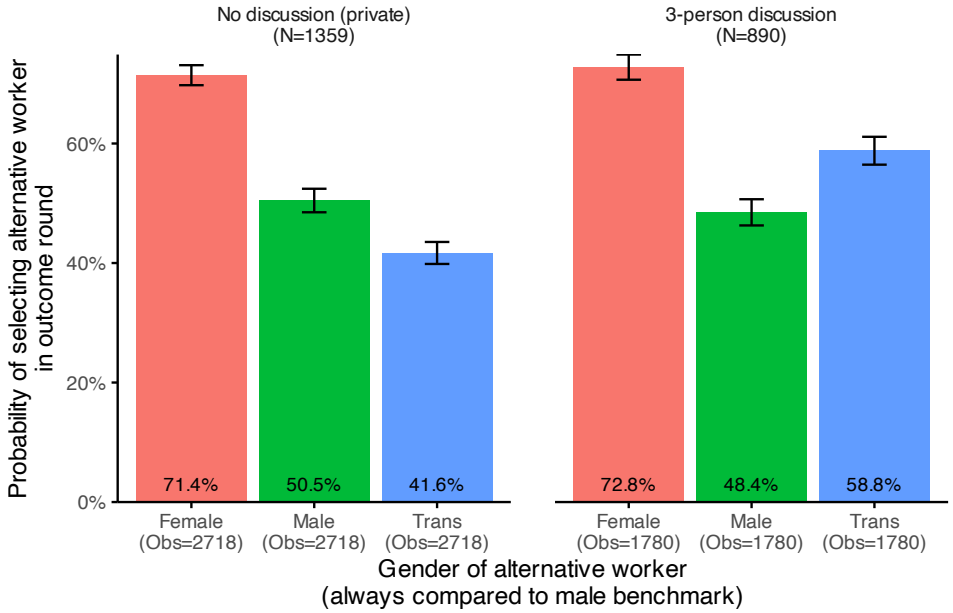
Notes: Unit of observation is participant × choice. Outcome is whether the participant selects the alternative worker instead of the male benchmark worker. An option (P) (weakly) dominates an option (Q) if it is strictly better on at least one characteristic, and is not worse on any characteristic. More specifically, P weakly dominates Q when (i) P either offers more items than Q, or P has a higher reliability score than Q (if it is shown), or both; *and* (ii) Q does not offer more items than P, *and* (iii) Q does not have a higher reliability score than P (if it is shown). *Trans dominates* is when the transgender worker weakly dominates the other option. *Trans is dominated* is when the transgender worker is weakly dominated by the other option. *Neither dominates* is when neither the transgender worker nor the other option dominates. Controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; phase fixed effects; whether the alternative worker was shown on the right; and the controls selected by double LASSO (see Section **??**).

Figure A11: *Predictions about group members' private choices depends on positivity of discussion*



Notes: *Notes*: *P(Other group members spoke trans)* (x-axis) is the proportion of (discussion choice × participants) that spoke positively about a transgender option. There were two other group members, each with two opportunities to speak positively about a transgender worker. *P(Predicted group would choose trans)* (y-axis) is the predicted probability that a fellow group member would choose the transgender worker in the private outcome round. Plotted line of best fit and 95% confidence intervals are based on raw data with no controls. Left panel uses 3-person discussion participants in phases 1 and 2; right panel uses listeners in phase 2. Slope and p-values come from models with LASSO controls and all other controls in **??**. β=0.36 implies that each additional positive discussion from a group member corresponds to being $36/4 = 9$ p.p. more likely to select a transgender worker in the outcome round. H_0(*intercept=control*) tests whether a discussion participant for whom *P(Other group members spoke trans) = 0* exhibits different post-discussion discrimination from the control group.

Figure A12: *Probability of selecting the alternative worker for each gender separately*



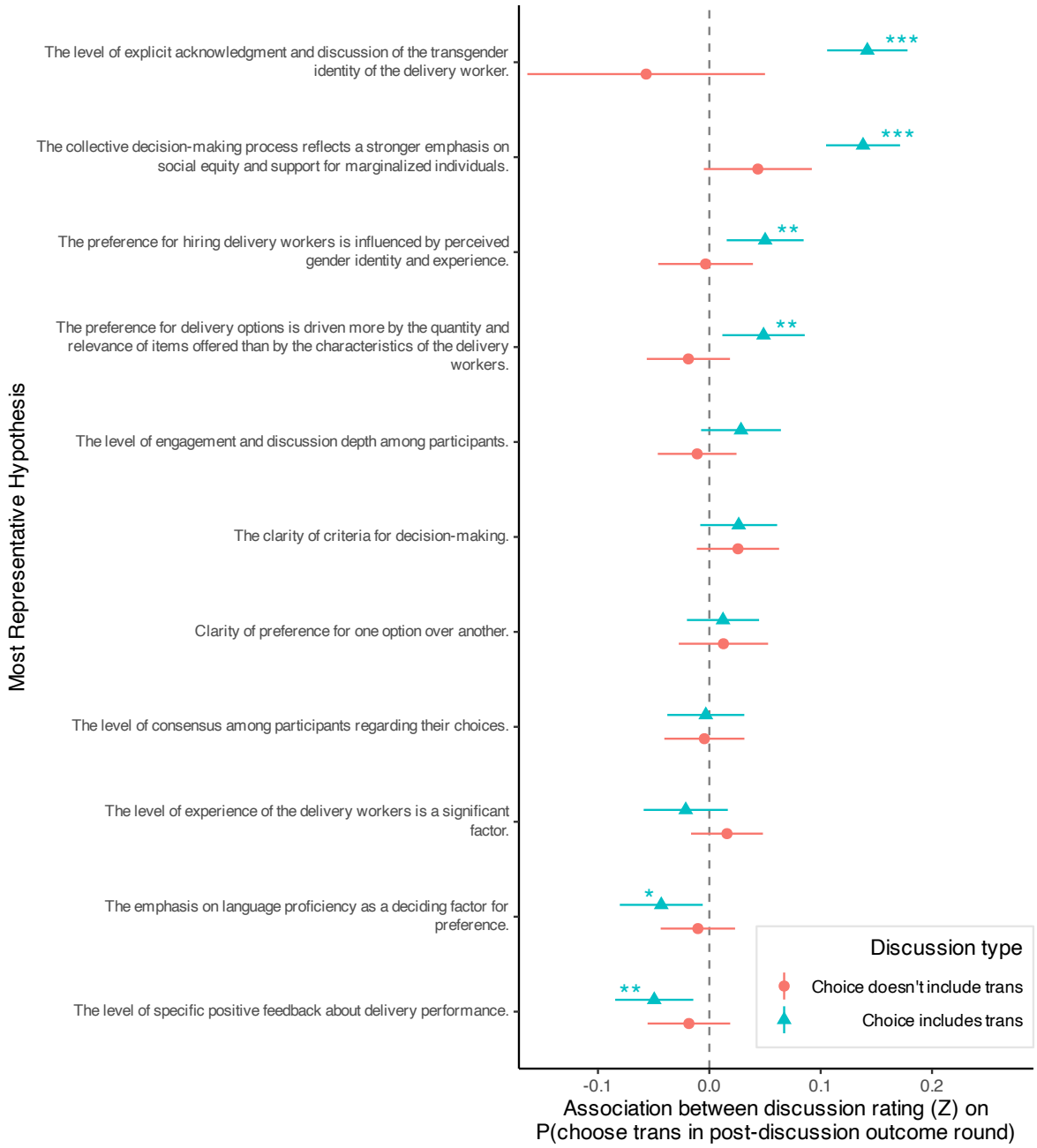
Notes: The unit of observation is participant × choice. The sample includes participants in the *No discussion (private)* and *3-person discussion* round. Only choices from the private outcome round are included. The outcome is whether the participant selected the *alternative worker*, who could be male, female, or transgender, instead of the male benchmark worker. Each participant saw two choices where the alternative worker was female, two choices where the alternative worker was male, and two choices where the alternative worker was transgender.

Table A13: *Sentence-level transcript analysis using k-means clustering*

		% sentences in discussion		Association between % sentences in trans discussions and P(chose trans in outcome round)	
ID	Representative quotes	No trans	Includes trans	β (p.p.)	q-value
1	'We should give them the opportunity.'; 'We should give them an opportunity.'; 'We should give them opportunities.'	1.3	7.1	6.8	0.00***
2	'The person is transgender.'; 'The transgender.'; 'For transgender.'	0.3	11.5	5.7	0.00***
3	'He looks good.'; 'He appears decent.'; 'He doesn't seem confident. He looks good to me.'	8.4	4.1	-7.3	0.02**
4	'Right.'; 'Alright, say.'; 'Okay.'	17.4	15.8	-4.0	0.06*
5	'They are doing the delivery.'; 'They have done many deliveries'; 'They have made more deliveries.'	5.9	3.4	-4.1	0.12
6	'For them.'; 'They will be there for work, like that'; 'Let them work as well.'	5.0	10.6	2.6	0.34
7	'B.'; 'B ..'; 'b'	3.4	4.5	1.9	0.48
8	'A, A.'; 'Yes, A only'; 'Let's go with A.'	7.4	7.2	-1.9	0.68
9	'This has chili powder, tea powder, this has chili powder.'; 'Chili powder, tea powder'; 'They have given chili powder, tea powder.'	3.3	1.4	-1.6	0.68
10	'They know Tamil and English.'; 'They know both Tamil and English, that's why.'; 'He knows both Tamil and English.'	3.6	2.6	-1.3	0.68
11	'Has got ten out of ten.'; 'They scored 8 out of 10.'; 'He scores ten out of ten.'	4.0	2.7	-1.7	0.75
12	'Both are the same items.'; 'Both have the same item'; 'Items are the same.'	6.9	3.9	-1.3	0.75
13	'They have 5 years of experience. He only has 4 years.'; 'He has 5 years of experience, this one has 4 years.'; 'This one has 4 years of experience, the other has 5 years.'	3.1	2.0	-1.8	0.80
14	'No'	1.2	1.2	0.3	1.00
15	'Yes.'	3.3	3.2	0.9	1.00
16	'Same reason.'; 'That's the reason.'; 'The same reason.'	1.5	2.0	-0.3	1.00
17	'a'; 'A.'	4.4	4.1	0.6	1.00
18	'Only ladies.'; 'It should be ladies.'; 'There are ladies.'	7.8	1.8	-0.5	1.00
19	'It's B only.'; 'What about B?'; 'Also B,'	6.7	7.6	-0.3	1.00
20	'Should we choose someone?'; 'Which person should we choose?'; 'Do we have to choose who?'	5.1	3.0	0.1	1.00

Notes: All sentences in the transcripts are clustered into 20 semantically similar clusters using OpenAI's embeddings and k-means clustering (Appendix **??**). Each row represents a cluster. 3 representative quotes are chosen that have the greatest cosine similarity to the cluster mean. *% sentences in the discussion* denotes the proportion of sentences that belong to that cluster, for discussions without a transgender option (column 3), and for discussions with a transgender option (column 4). Column 5 is the coefficient of a regression of the group-level proportion of utterances in that cluster for transgender conversations, with the outcome as *P(chose trans in outcome round)*. For this regression the proportion of utterances is standardized so it can be interpreted in standard deviations (e.g., if a group is 1 SD more likely to utter statements in cluster 1, this is associated with a 6.8 p.p. increase in the probability of selecting transgender workers). q-values are the **?** multiple-hypothesis-adjusted p-values. Rows are ordered by ascending q-values. Correlation with non-trans quotes are all non-significant (all q-values=1).

Figure A14: *Transcript-level features that predict reductions in discrimination*



Notes: AI generated 500 hypothesis about the differences between random pairs of transcripts, and then rated every transcript according to how much it fulfilled that hypothesis on a scale of 1-10 (see Appendix **??**). I reduce the dimensionality of the 500 hypothesis using factor analysis to 11 factors. Each hypothesis "type" is represented by the specific hypothesis that loads most highly on that type (the labels in the y-axis). I take the average rating at the group level for both discussions without a transgender option and with a transgender option, and then regress these on the probabilty of selecting a transgender worker in the outcome round. There is one regression for each hypothesis type. Coefficients and 95% CIs are shown for both regressors. All regressions include all main controls from column 3 of **??**. * $q < 0.1$, ** $q < 0.05$, *** $q < 0.01$, where q are the multiple-hypothesis-adjusted p-values, taking all 40 coefficients together as a set for adjustment (**?**).

Table A15: *No evidence of differential attrition*

	Dep var: Follow-up survey completed (=1)	
	3-person discussion sample (Phase 1 + 2)	Rights videos (all participants)
	(1)	(2)
3-person discussion	−0.001	
	(0.010) [0.881]	
Rights messaging video		0.000
		(0.010) [1.000]
Legal rights video		0.005
		(0.010) [0.581]
Num. observations	2249	3397
Num. participants	2249	3397
Num. groups	751	1134
Outcome mean: No discuss. (priv.)	0.96	
Outcome mean: Control vid.		0.95

Notes: * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant. Dependent variable is whether the follow-up survey was completed. Column (1) includes only participants in the *No discussion (private)* or the *3-person discussion* arms, in both phases. Column (2) includes all participants in phase 2. Column (3) includes all participants in both phases. includes choices that involved a transgender worker.

Table A16: *Treatment round choices (3-person discussion sample, Phases 1 and 2)*

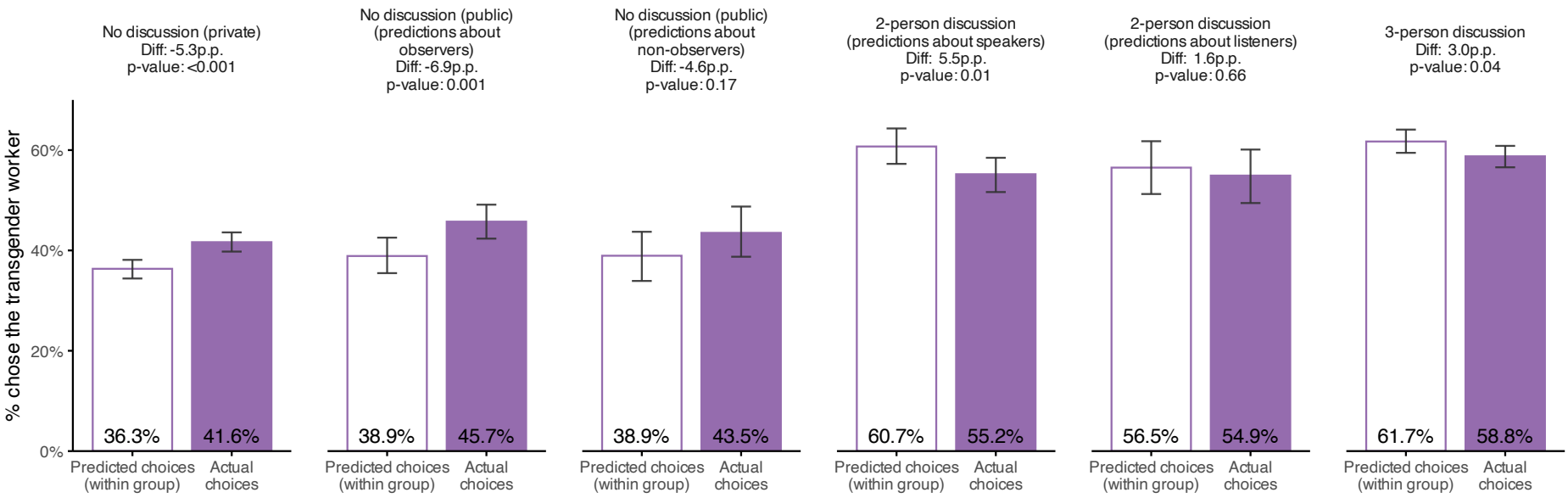
	Chose worker in treatment round (=1)		Chose trans in treatment round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans × 3-person discussion	0.198***	0.197***	
	(0.031) [<0.001]	(2.9×10^{-2}) [<0.001]	
Worker is trans	−0.085***		
	(0.015) [<0.001]		
3-person discussion	0.007	0.004	0.199***
	(0.023) [0.720]	(2.1×10^{-2}) [0.829]	(0.022) [<0.001]
Num. observations	8996	8996	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Outcome mean†: no discuss (priv.), worker non-trans	0.60	0.60	
Outcome mean: no discuss (priv.), worker trans	0.51	0.51	0.51
Controls		X	X
Controls interacted with worker is trans		X	

Notes: * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The outcomes are based on *treatment round choices*, i.e., during the discussion in the 3-person discussion arm. The specification used is seen in equation **??**, and is otherwise the same as Tables **??** and **??**.

Figure A17: *Individual-level estimates of willingness to pay to choose a transgender worker in the outcome round*

Notes: Plots the distribution of individual-level estimates of willingness to pay (WTP) to select a transgender worker. Estimates come from a version of Equation **??** that includes an interaction of individual fixed effects γ_i and $Trans_{ijk}$. WTP is calculated in monetary terms by dividing γ_i by the coefficient on relative item value (in Rupees) across the two options. Estimates cluster around 3 peaks corresponding to choosing a transgender worker 2/2 times, 1/2 times, or 0/2 times. Variation between these peaks is driven by variation in the items offered and the characteristics of the worker (e.g., experience, reliability score).

Figure A18: *Predictions about others in group*



Notes: Sample includes all participants. Unit of observation is participant × prediction. Only choices that include a transgender photo are included. Hollow bars represent the probability that a participant predicts that their group member selects a transgender delivery worker. The prediction was incentivized. Each participant made 2 predictions (one involving a transgender worker) for each of their 2 group members. The two predictions involving a transgender worker are included for analysis. Filled bars represent the actual probability that participants select a transgender worker in the outcome round (restricting to only choices for which another group member made a prediction). *2-person discussion (predictions about speakers)* includes all predictions made *about* the private choices of the speakers in the discussion. *2-person discussion (predictions about listeners)* includes all predictions made *about* the private choices of the people who just listened to the discussion.

Table A19: *Mediation analysis of discussion using mechanism outcomes*

	Dep var: Chose trans in private outcome round				
	(1)	(2)	(3)	(4)	(5)
3-person discussion	0.155***	0.068***	0.165***	0.163***	0.075***
	(0.020)	(0.018)	(0.020)	(0.019)	(1.7×10^{-2})
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Predicted prob. of others choosing trans (community)	0.329***				0.111***
	(0.031)				(3.0×10^{-2})
	[<0.001]				[<0.001]
Predicted prob. of others choosing trans (group)		0.413***			0.349***
		(0.019)			(2.1×10^{-2})
		[<0.001]			[<0.001]
Disagreed with discrimination (=1)			0.184***		0.083**
			(0.042)		(3.9×10^{-2})
			[<0.001]		[0.033]
Trans likely to complete delivery (=1)				0.229***	0.149***
				(0.018)	(1.7×10^{-2})
				[<0.001]	[<0.001]
Num. observations	4498	4476	4498	4498	4476
Num. participants	2249	2249	2249	2249	2249
Num. groups	751	751	751	751	751
Outcome mean: No discuss. (priv.)	0.42	0.42	0.42	0.42	0.42

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. Only choices involving a transgender worker are included. The dependent variable is whether the transgender worker was selected in the private outcome round choices. Additional variables are based on the mechanism outcomes described in Appendix **??**. Controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; relative reliability score; relative items offered; whether the reliability score was shown; and the controls selected by double LASSO (see Section **??**).

Table A20: *Effects on hypothetical choices after 2–9 weeks for mechanism treatments*

	Chose trans in follow-up round (=1) (pairs with trans only)
	(1)
3-person discussion	0.074**
	(0.031) [0.016]
Speaker (2-person discussion)	0.088***
	(0.030) [0.004]
Listener (2-person discussion)	0.114***
	(0.034) [<0.001]
No discussion (public)	−0.002
	(0.025) [0.949]
Num. observations	4160
Num. participants	2080
Num. groups	738
Outcome mean†: no discuss (priv.), worker trans	0.46
p(3-person discussion=Speaker (2-person discussion))	0.714
p(3-person discussion=Listener (2-person discussion))	0.324
p(3-person discussion=No discussion (public))	0.022
p(Speaker (2-person discussion)=Listener (2-person discussion))	0.460
p(Speaker (2-person discussion)=No discussion (public))	0.006
p(Listener (2-person discussion)=No discussion (public))	0.001

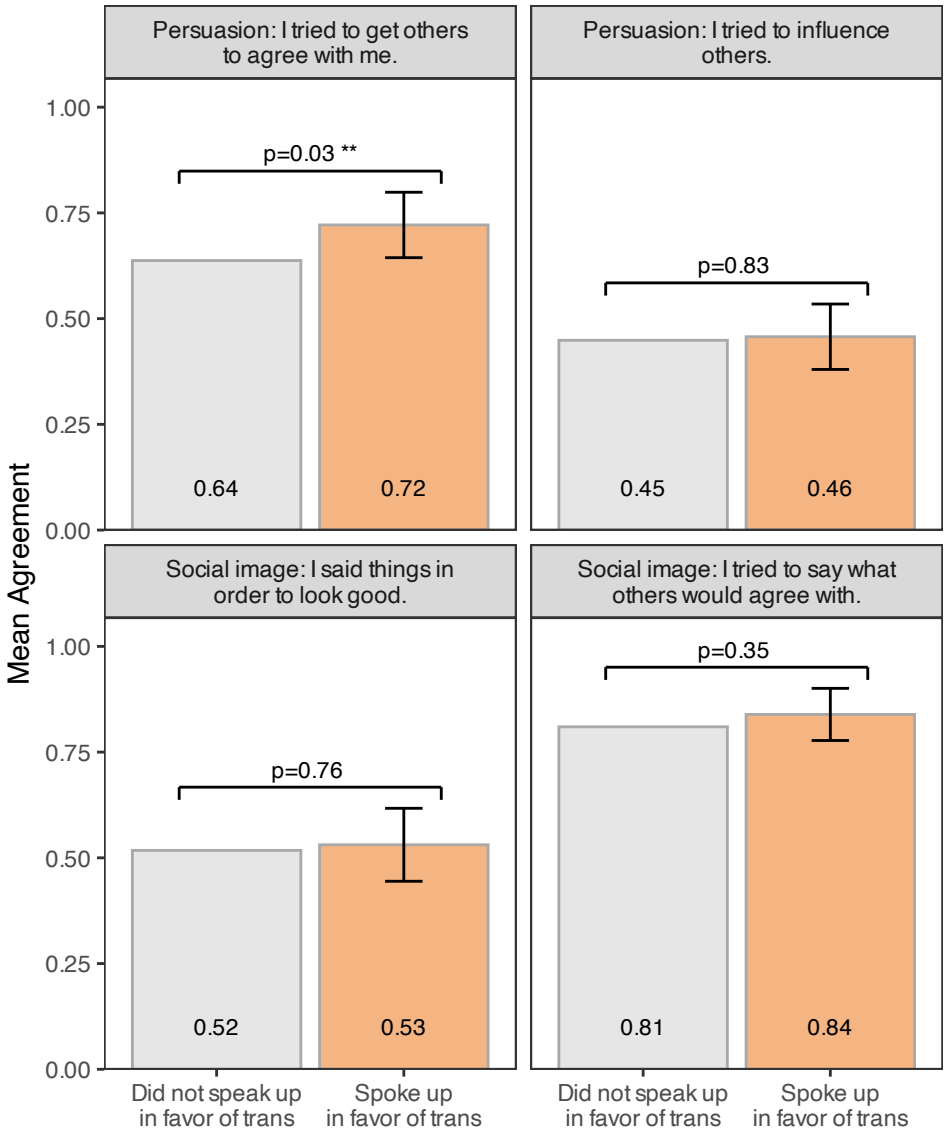
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all participants in phase 2. Outcome is whether the transgender worker was chosen; only pairs involving a transgender worker are included. Controls include stratum fixed effects; dummies for the discussion-arm treatments; whether the alternative worker was shown on the right; phase fixed effects; and the controls selected by double LASSO (see Section **??**).

Table A21: *Correlation between dominance in discussion and post-discussion pro-trans choices (3-person discussion arm only)*

	Dep var: Chose trans in private outcome round (=1)		
	Combined index (Z)	Spoke first	Was dominant
	(1)	(2)	(3)
Dominance index (Z)	−0.033**		
	(0.016) [0.037]		
Dominance index - transgender choices only (Z)	0.048***		
	(0.015) [0.002]		
P(spoke first)		−0.098	
		(0.064) [0.125]	
P(spoke first) - transgender choices only		0.107**	
		(0.048) [0.027]	
P(dominated conversation)			−0.099
			(0.074) [0.184]
P(dominated conversation) - transgender choices only			0.149**
			(0.063) [0.019]
Num. observations	1776	1776	1776
Num. participants	890	890	890
Num. groups	297	297	297

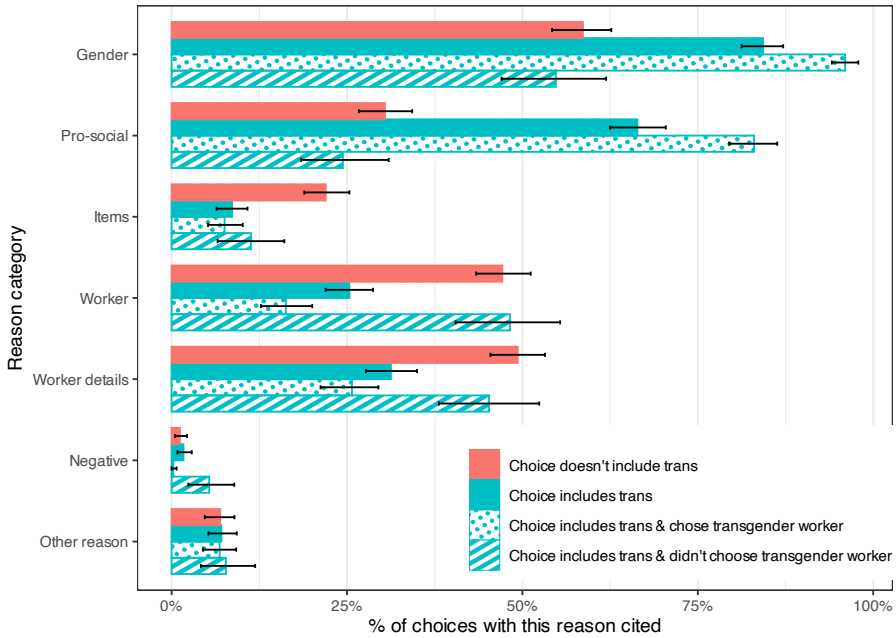
Notes: *P(spoke first)* is the probability that a participant spoke first in their group in the discussion of a choice, as marked by enumerator observations. The mean is 33%. *P(dominated)* is the probability that a participant dominated the discussion of a choice, as marked by enumerator observations. The mean is 0.55 (more than one participant can be marked as dominating). *Dominance index (Z)* is the sum of normalized (Z-index) values for *P(spoke first)* and *P(dominated)*. Only *3-person discussion* arm is included. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets. Unit of observation is the participant $\times$ choice level. Outcome is whether the transgender worker was selected in the private outcome round (i.e., after the discussion). Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; the relative # items offered by the alternative worker, the relative reliability score of the worker, and a dummy for whether the reliability score was shown.

Figure A22: *Motivations during the discussion*



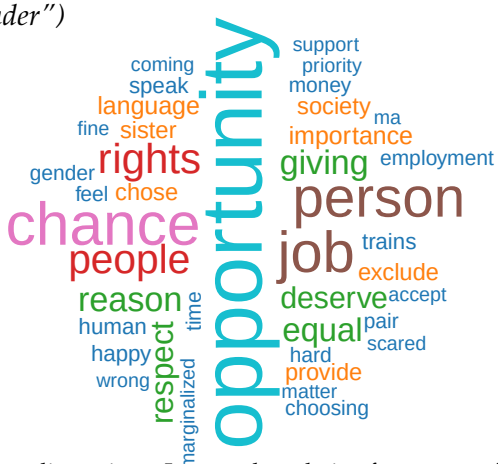
Notes: 51% of participants in the discussion spoke in favor of trans workers for both choices. p-value is calculated using a regression controlling for stratum fixed and rights videos, and clustering standard errors at the group level. Phase 1 only, 3-person discussion participants, N=576.

Figure A23: *Reasons cited in the 3-person discussions (enumerator observations)*



Notes: Unit of observation is a group × choice. Sample is the *3-person discussion* arm in both phase 1 and 2. Confidence intervals are based on a bootstrapped binomial distribution. One enumerator observed the discussion and marked the main reasons that the participants said they were selecting the chosen option during the discussion. *Gender* includes saying that the worker is transgender, male, or female. *Pro-social* reasons include (i) wanting to give an opportunity or help the worker, (ii) saying that the worker is also human, (iii) saying that the chosen worker seems poor, (iv) saying "We shouldn't discriminate". *Items* is when participants say they chose the option because it offered more items. *Worker* includes saying (i) it would be easy to talk with the worker, (ii) the choice is based on how the worker looks / the photo, (iii) the worker seeming reliable, (iv) the worker seeming friendly, (v) it being easy to relate to the worker, (vi) the perceived age of the worker. *Worker details* includes reasons based on written details on the worker profile: (i) the reliability score, (ii) whether they speak English, (iii) their experience, or (iv) their education. *Negative* is when the reason cited is a negative comment about the worker that was not chosen (e.g., the other person looks scary or indecent).

Figure A24: *Word cloud: most relatively common words in discussions about transgender workers (excluding the word "transgender")*



Notes: Using transcript data from discussions, I count the relative frequency of all words used by participants. I exclude stop words such as "a" or "the". The cloud shows words for which the *difference* in relative frequency when discussing transgender workers vs non-transgender workers is highest. The word "transgender" scores highest on this measure but is excluded.

Table A25: *No detectable heterogeneity by discussant persuasiveness or group relations*

	Chose trans in private outcome round (Phase 2 only)			
	No discussion (private) + listeners		No discussion (private) + 3-person discussion	
	(1)	(2)	(3)	(4)
Listened to 2-person discussion	0.126***	0.156***		
	(0.048)	(0.049)		
	[0.009]	[0.002]		
3-person discussion			0.180***	0.160***
			(0.040)	(0.041)
			[<0.001]	[<0.001]
High persuasiveness score for discussants	−0.016		−0.016	
	(0.027)		(0.027)	
	[0.549]		[0.554]	
Listened to 2-person discussion × High persuasiveness score for discussants	0.033			
	(0.066)			
	[0.615]			
3-person discussion × High persuasiveness score for discussants			0.006	
			(0.051)	
			[0.907]	
Close relations with others in group		0.028		0.027
		(0.029)		(0.029)
		[0.347]		[0.359]
Listened to 2-person discussion × Close relations with others in group		−0.028		
		(0.068)		
		[0.677]		
3-person discussion × Close relations with others in group				0.055
				(0.055)
				[0.325]
Num. observations	1878	1878	2140	2140
Num. participants	939	939	1070	1070
Num. groups	436	436	358	358
Controls	X	X	X	X

Notes: *High persuasiveness score for discussants*: Above median score for the other two participants in a group on an index of persuasiveness. Index is constructed using a weighted sum of the ratings out of 10 given for the following character traits of other participants: (i) confident; (ii) quiet; (iii) like a leader; (iv) shy; (v) talkative; (vi) admirable; (vii) inspiring. See Section **??** for details.

Close relations with others in group: Above median score on an index of perceived relationships with other participants in the group (see section **??** for full details). The index is constructed using a weighted sum of (i) whether the other participant is a close family member, (ii) another family member, (iii) a friend, or (iv) simply a neighbor; (v) how long they have known the other participant; (vi) how often they talk to the other participant; (vii) how often they ask the other participant for advice; (viii) how often they ask for recommendations for what to buy; (ix) how often they tell secrets to the other participant. For each participant, I take the mean score of their ratings for the two other participants in their group to get a score at the participant level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample in columns 1-2 includes only *No discussion (private)* and *listeners* in the *2-person discussion* arm. Sample in columns 3-4 includes only *No discussion (private)* and *3-person discussion* arms. Only phase 2 of data collection is included (when group relationships were elicited). The outcome is whether the transgender worker was selected in the private outcome round, restricting analysis to only choices that include a transgender worker. Additional controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; relative # items offered; relative reliability score; whether the reliability score was shown.

B Experimental design: further detail

Item randomization. The number of grocery items on offer was set so that within a pair, both options had an equal number of items 60% of the time, one option had one extra item 30% of the time, and one option had two extra items 10% of the time.

Reliability score. Some choice-pairs reported the true proportion of successful deliveries from a set of timed training exercises carried out by all workers (the "reliability score"). This reliability score incorporated exogenous variation in the perceived quality of each worker. Participants were told that this was the proportion of completed deliveries from a training exercise. Workers completed multiple training exercises with different time limits, and I randomly showed their score within one of three categories: their low score (5 or 6), their mid-value score (7 or 8), or their high score (9 or 10). (see **??** for discussion of the ethical considerations).

Other signals of quality. In addition, for some choice-pairs, I truthfully reported (i) whether workers had 0-4 years or 5 years or more of work experience, and (ii) whether the worker spoke both Tamil and English or just Tamil. I sampled photos so that these characteristics were balanced across each worker gender.

Treatment-round protocols for mechanism treatments. To make social image concerns salient, *No discussion (public)* participants chose in a group setting. To ensure participants did not influence each other *during* the elicitation process, they were told not to show others their choices and to remain silent. Participants saw others' responses only 1.6% of the time, and someone commented on a delivery option in only 6.0% of groups. All three participants in a given group always saw the same delivery options in the treatment round, regardless of treatment status. 93% listeners stayed silent about the choices.

Privacy in outcome round. The main results are robust to to dropping the 6% of cases where the outcome round was overheard by neighbors (**??**).

Video attention and comprehension. To ensure participants could hear the video and were concentrating fully, participants always watched the video alone using headphones, rather than in a group. All participants in a group-of-3 watched the same video, but were not told explicitly that others had seen the same video. After watching the video, they were asked comprehension questions about the content (and were corrected if they did not answer correctly), and then read the script of the video text again for 2 minutes.

Table B26: *Robustness to protocol fidelity*

	Dep var: Chose trans in private outcome round (=1)		
	Drop when others heard outcome-round answers (3-person discussion sample)	Drop when listener spoke (Phase 2 sample)	Drop when No-discussion (public) participants spoke (Phase 2 sample)
	(1)	(2)	(3)
3-person discussion	0.174***	0.178***	0.179***
	(2.1×10^{-2})	(0.031)	(0.031)
	[<0.001]	[<0.001]	[<0.001]
Observer (No discussion, public)		0.046*	0.042
		(0.026)	(0.027)
		[0.080]	[0.112]
Non-observer (No discussion, public)		0.024	0.028
		(0.031)	(0.032)
		[0.436]	[0.388]
Speaker (2-person discussion)		0.141***	0.134***
		(0.029)	(0.028)
		[<0.001]	[<0.001]
Listener (2-person discussion)		0.144***	0.127***
		(0.034)	(0.033)
		[<0.001]	[<0.001]
Num. observations	4178	4364	4364
Num. participants	2089	2182	2182
Num. groups	750	729	729

Notes: Sample in column 1 includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2, but excludes cases where the respondent said that others could hear their private outcome-round responses. Column 2 is the phase 2 sample, but excludes the cases when the listener spoke during the 2-person discussion. Column 3 is the phase 2 sample, but excludes cases when any of the *No discussion (public)* participants spoke during the treatment round, which was supposed to be silent. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant $\times$ choice level. Only choices involving a transgender worker are included. The dependent variable is whether the transgender worker was selected in the private outcome round choices. Controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; relative reliability score; relative items offered; whether the reliability score was shown; and the controls selected by double LASSO (see Section **??**).

C Discussion design details

To encourage people to speak up in the discussions, the surveyor leading the discussion asked icebreaker questions before the treatment round started. In the *No discussion (public)* arm, participants sat together in a group and also took part in this icebreaker activity. *No discussion (private)* participants were asked the same icebreaker questions, but individually and in private.

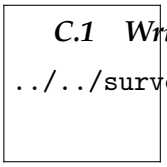
To encourage discussion about a number of different characteristics in the treatment round, 2 out of the 4 choice-pairs in the treatment round included information about experience and language for both workers, and all choice-pairs included the reliability score for both workers.

The enumerator who led the discussion was told to prompt participants to speak using neutral questions that did not lead the participants to prefer one option or the other (for example, "What are the differences between A and B?"). They were also explicitly told never to use the word "transgender" themselves, in order to avoid revealing the purpose of the experiment to the participants.

For the 2-person discussion, the enumerator leading the discussion also asked the listener if he or she heard the choice being made by the speakers, along with the reason given by the speakers. If the listener did not hear, the speakers were asked to repeat themselves.

The discussion script used by the enumerator leading the discussion is below (Section **??**).

C.1 Written discussion script for facilitator



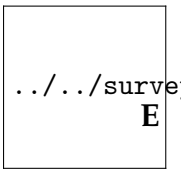
D Ethical considerations

The study aimed to obtain a revealed-preference measure of discrimination based on real-stakes choices. Using real rather than hypothetical choices was crucial for reducing concerns social desirability bias and experimenter demand effects, and therefore for understanding methods for reducing such discrimination. However, this approach meant balancing multiple ethical considerations—primarily avoiding explicit deception of respondents while also protecting transgender workers from harm.

An important concern is that participants might view the protocol as deceptive because they were unlikely to receive a delivery from a transgender worker. This concern had to be weighed against the risks faced by transgender workers performing deliveries. These workers could be exposed to stigma and abuse when visiting participants' homes. The randomization aimed to minimize this risk while truthfully telling participants they could receive a delivery from any chosen worker. For the few transgender workers who performed deliveries, each was accompanied by 2-3 enumerators. Interaction was minimized, and these enumerators were trained to avoid conflict and protect the worker. This design protected transgender workers while maintaining truthful revealed preference responses.

A second concern is that participants lacked sufficient context to properly interpret the reliability score shown on some worker profiles, though the score itself was truthful. Using the reliability score was important for examining whether discrimination against transgender workers was statistical. The reliability score truthfully reported how many deliveries a worker successfully completed in a training exercise. However, participants were unaware that workers undertook multiple training exercises of different durations, yielding different scores. Participants were told (e.g.) "8/10 means that out of 10 deliveries they had to make in a training exercise, 8 times they successfully delivered." All information provided was true, without additional false details about the exercise's nature: they weren't told that the exercise was conducted only once. This mimics real-world situations, where employers often have incomplete information about job candidates.

Finally, participants' primary concern about reliability was ensuring delivery receipt. In practice, we completed deliveries to 95.7% of them. Participants were therefore always selecting an extremely reliable delivery worker, so any misleading inference from the reliability score did not materially affect them.



E Video scripts

../../survey_materials/video_scripts.pdf

F Pre-analysis plan

The study was pre-registered in the AEA registry under the ID # AEARCTR-0010953. Two pre-analysis plans were uploaded: the first in March 2023, corresponding to the start of phase 1, and the second in May 2023, corresponding to the start of phase 2. In phase 1, I faced data quality issues and unexpectedly low survey productivity in the first 2 days of data collection. This, along with a tight budget, meant that I decided to cut the sample size and the survey length, resulting in design changes relative to the phase 1 pre-analysis plan. As noted in the main text, phase 2 of data collection was added to the design upon the receipt of additional funding in the course of the experiment, resulting in the updated pre-analysis plan. I outline all the deviations from the pre-analysis plan, along with their justifications, below.

F.1 Phase 1

- *Mixed-video arm*. In phase 1, I had planned to include 450 individuals in a "mixed-video" arm. Because of budget constraints and low productivity, I decided at the start of phase 1 to remove this treatment condition, reducing the planned sample size. Because of this, I also dropped the plan to analyze spillover effects between individuals in a group.
- *High-stakes condition*. In phase 1, I had planned to randomize half of every treatment group into the "high-stakes condition" (i.e., for them to receive 3 deliveries instead of 1). However, because this tripled the expenditures on grocery items, I decided to restrict the randomization to only a subsample of approximately 200 groups, half of whom were allocated to the high-stakes condition.
- *Attitude questions*. Participants' understanding of the measure of attitudes ("Disapproval of discrimination") that I had planned to use appeared to be poor, so I replaced it with a simpler self-reported attitude question.
- *Other mechanism questions*. In order to reduce the length of the survey, I also dropped some secondary mechanism measures, including: (i) an implicit association test; (ii) whether discrimination can lead to legal consequences; (iii) the perceived similarity index; (iv) some controls, including the number of children in a household, smartphone ownership, a measure of willingness to persuade in discussions, and a proxy for baseline progressive social attitudes.
- *Restriction to urban*. In order to reduce transport costs, I carried out surveys only in urban areas, instead of both rural and urban areas.

F.2 Updates for phase 2

- The *No discussion (public)* and *2-person discussion* arms were added.
- *Additional mechanism outcomes*. To allow for further analysis of the mechanisms behind the group discussion, I added measures of (i) relationships between group members, (ii) persuasiveness of group members, (iii) private grocery pick-up choices, and (iv) memory checks (i.e., how well do participants remember their own and others' choices).

- *Removed mechanism outcomes*. To avoid the survey becoming too long, I removed the measure of salience and the measure of social desirability score for phase 2.

F.3 Other changes

- *Delivery time*. I originally planned to carry out follow-up surveys and deliveries in parallel to the main surveys. However, it became clear that this was logistically infeasible, so I instead chose to carry out all deliveries at the end of each phase. This meant that the delivery time was 2–9 weeks, instead of the pre-specified 1 week.
- *Discussion recordings*. I planned to use discussion transcripts to encode a "Probability of endorsing" variable, separately for each individual. However, it became clear that it was not feasible for enumerators to accurately attribute each statement to a specific participant, so I do not include this for analysis.

F.4 Pre-specified analyses

Here I describe analyses that I specified in the pre-analysis plan, but which are not presented as main results in the text.

- *Video and discussion interactions*. The fully interacted specification that includes all video arms and the 3-person discussion arm variation was pre-specified and is shown in **??**.
- *Pooled phase 2 results*. In the phase 2 pre-analysis plan, I described that I would pool some treatment arms (see Figure 1 in the pre-analysis plan). As prespecified, I pooled the 2-person discussion and 3-person discussion participants when analyzing the treatment round (**??**). However, for reader clarity in the main text I did not pool any treatment arms when presenting the phase 2 outcome round results (**??**). The corresponding pooled results are presented in **??**.
- *Heterogeneity with respect to round 1 observations*. **??** shows the heterogeneous effects of *observers*' choices with respect to the round 1 choices they observed. **??** shows the heterogeneous effects of *listeners*' choices with respect to the round 1 choices they listened to.
- *Heterogeneity with respect to group composition*. **??** shows heterogeneous effects of the discussion with respect to persuasiveness and group relations. I find no detectable heterogeneity.

Table F1: *Effect of phase 2 treatments on private choices in outcome round (pooled)*

	Chose worker in private outcome round (=1)		Chose trans in private outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.200***		
	(0.016) [<0.001]		
Observer (No discussion, public)	0.000	−0.002	0.038
	(0.015) [0.998]	(0.014) [0.865]	(0.024) [0.117]
Listener (2-person discussion)	0.012	0.003	0.120***
	(0.020) [0.535]	(0.019) [0.856]	(0.032) [<0.001]
Discussion (pooled)	0.004	0.008	0.147***
	(0.013) [0.730]	(0.012) [0.494]	(0.022) [<0.001]
Worker is trans × Observer (No discussion, public)	0.045*	0.042	
	(0.027) [0.096]	(0.026) [0.101]	
Worker is trans × Listener (2-person discussion)	0.124***	0.125***	
	(0.039) [0.001]	(0.038) [0.001]	
Worker is trans × Discussion (pooled)	0.154***	0.145***	
	(0.025) [<0.001]	(0.024) [<0.001]	
Num. observations	13 308	13 308	4436
Num. participants	2218	2218	2218
Num. groups	741	741	741
Controls		X	X
Controls interacted with worker is trans		X	
p(Observer=Listener)	0.060	0.045	0.020
p(Observer=Discussion)	0.000	0.000	0.000
p(Listener=Discussion)	0.430	0.595	0.409

Notes: In this specification, I pool the *No discussion (private)* and the *No discussion, public (non-observers)*. They are the omitted category. I also pool *2-person discussion (speakers)* and *3-person discussion* participants, calling them *Discussion (pooled)*. * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all treatment arms in phase 2 of data collection. Column (3) only includes choices that involved a transgender worker. In columns (1) and (2), the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was selected in the private choices in the *outcome round*. In column (3), it is whether the transgender worker was selected. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. The specification used is seen in equation **??**. Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; and the controls selected by double LASSO (see Section **??**). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown. Columns (2) and (3) also include controls for the relative # items offered by the alternative worker, the relative reliability score of the worker, and a dummy for whether the reliability score was shown. Randomization inference *p*-values at the base of the table test for differences between treatment effects across treatment arms, i.e., for differences in the interacted terms in columns (1) and (2), and differences in the uninteracted terms in column (3).

Table F2: *Effect of observing others' choices*

	Dep var: Chose transgender worker in private outcome round (=1) (Phase 2 only)					
	Sample: No discussion (private)	Sample: Non-observers	Sample: Observers	Sample: No discussion (private) + Non-observers	Sample: No discussion (private) + Observers	Sample: Non-observers +Observers
	(1)	(2)	(3)	(4)	(5)	(6)
π_{-i} = P(others in group selected trans in treatment round)	−0.036	0.123	0.208***	−0.035	−0.040	0.116
	(0.052) [0.494]	(0.086) [0.156]	(0.065) [0.002]	(0.052) [0.503]	(0.051) [0.431]	(0.081) [0.156]
P(selected trans in treatment round)	0.384***	0.370***	0.427***	0.389***	0.399***	0.412***
	(0.033) [<0.001]	(0.067) [<0.001]	(0.051) [<0.001]	(0.029) [<0.001]	(0.028) [<0.001]	(0.040) [<0.001]
Non-observer (No discussion, public)				−0.057		
				(0.055) [0.307]		
π_{-i} × Non-observer (No discussion, public)				0.159		
				(0.099) [0.108]		
Observer (No discussion, public)					−0.098**	−0.033
					(0.043) [0.024]	(0.058) [0.562]
π_{-i} × Observer (No discussion, public)					0.272***	0.095
					(0.079) [<0.001]	(0.102) [0.354]
Num. observations	1512	398	798	1910	2310	1196
Num. participants	756	200	399	956	1155	599
Num. groups	253	200	200	453	453	200
Controls	X	X	X	X	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. *p*-values are in brackets. For coefficients involving randomized treatments, they are calculated using randomization inference. Unit of observation is the participant × choice level. The outcome in all columns is whether the participant chose a transgender worker in the private outcome round, restricting analysis to only choices involving a transgender worker. Sample only includes phase 2 of data collection. Column 1 only includes the *No discussion (private)* arm. Column 2 only includes the *Non-observers* from the *No discussion (public)* arm, who knew they were choosing publicly in the treatment round but did not observe others' choices before making outcome round choices. Column 3 only includes the *Observers* from the *No discussion (public)* arm, who were told others' choices before making their outcome round choices. Columns 4-6 include combinations of each of these treatment conditions. *P(others in group selected trans in treatment round)* (π_{-i}) is the proportion of times (out of a maximum of 4) that the other two participants in the group selected a transgender worker in the treatment round. *P(selected trans in treatment round)* is the proportion of times (out of a maximum of 2) that the participant herself selected a transgender worker in the treatment round. Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; relative # items offered by the transgender worker; relative reliability score; and a dummy for whether the reliability score was shown.

Table F3: *Effect of listening to a discussion that selected transgender workers*

	Dependent var: Chose trans in private outcome round (Phase 2 only)							
	Sample: No discussion (private)		Sample: No discussion (private) + Listeners		Sample: Non-observers + Listeners		Sample: No discussion (private) + Non-observers + Listeners	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
π_{-i} = P(others in group selected trans in treatment round)	−0.017	−0.036	0.050	0.007	0.207**	0.195**	0.050	0.003
	(0.058) [0.774]	(0.053) [0.492]	(0.057) [0.379]	(0.056) [0.905]	(0.089) [0.021]	(0.089) [0.030]	(0.057) [0.378]	(0.057) [0.954]
P(selected trans in treatment round)		0.383***						
		(0.034) [<0.001]						
Listened to 2-person discussion			−0.114*	−0.112*	−0.060	−0.045	−0.114*	−0.127*
			(0.066) [0.084]	(0.066) [0.090]	(0.077) [0.440]	(0.079) [0.571]	(0.066) [0.084]	(0.065) [0.052]
π_{-i} × Listened to 2-person discussion			0.360***	0.358***	0.203*	0.181	0.360***	0.373***
			(0.093) [<0.001]	(0.093) [<0.001]	(0.116) [0.080]	(0.116) [0.119]	(0.093) [<0.001]	(0.093) [<0.001]
Non-observer (No discussion, public)							−0.054	−0.076
							(0.060) [0.365]	(0.060) [0.205]
π_{-i} × Non-observer (No discussion, public)							0.157	0.186*
							(0.106) [0.139]	(0.105) [0.079]
Num. observations	1512	1512	1878	1878	764	764	2276	2276
Num. participants	756	756	939	939	383	383	1139	1139
Num. groups	253	253	436	436	383	383	636	636
LASSO controls				X		X		X
Other controls	X	X		X		X		X
p-value: π_{-i} x Listener = π_{-i} x Non-observer							0.079	0.105

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. *p*-values are in brackets. For coefficients involving randomized treatments, they are calculated using randomization inference. Unit of observation is the participant × choice level. The outcome in all columns is whether the participant chose a transgender worker in the private outcome round, restricting analysis to only choices involving a transgender worker. Sample only includes phase 2 of data collection. Columns 1-2 only include the *No discussion (private)* arm. Columns 3-4 only includes the *No discussion (private)* arm and the *Listeners* who watched and listened to the 2-person discussion. Columns 5-6 include only the Listeners and the *Non-observers*, who knew their choices in the treatment round would be public, but who weren't told the choices of others before making their outcome round choices. Columns 7-8 include the No-discussion (private) arm, the Non-observers, and the Listeners. *P(others in group selected trans in treatment round)* (π_{-i}) is the proportion of times that the other two participants in the group selected a transgender worker in the treatment round. In the case of listeners, this is out of a maximum of 2 (since the others in their group, the speakers, are make two joint choices for the choices involving transgender workers). In the case of the no-discussion (private) and non-observers, it is out of a maximum of 4, since other participants can make different choices. *Other controls* include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; relative # items offered by the transgender worker; relative reliability score; and a dummy for whether the reliability score was shown. *LASSO controls* are those selected by double LASSO (see Section **??**).

G Data and measurement

G.1 Predicted choices (community)

Participants first made incentivized predictions about the choices of others in the study whom they did not know. They were shown 3 pairs of delivery options and were truthfully informed that 20 other people in the study, from the participants' area, had been shown those same pairs. They had to predict how many of those 20 picked each option. If participants made the closest guess, on average, across all 3 pairs, they were entered into a lottery to win 3000 Rs.' worth of additional items. Two of the 3 pairs were male-to-male comparisons, and 1 pair compared a male and a transgender person. A randomly selected half of the participants were always asked how many picked the option on the left, and half were asked how many picked the option on the right. The phrasing used was: "*In your opinion, how many people out of 20 chose the person on the [right/left]?*" The transgender option always appeared on the side being asked about.

G.2 Predicted choices (own group)

Participants made incentivized predictions of the private hiring choices of the other two people in their group. For each of the other two group members, they were asked to predict which option they chose for two pairs of delivery options. For each other person, one choice-pair compared a male and a male, and another compared a male and a transgender person. If they correctly guessed all 4 combinations (2 predictions each of the 2 group members) they were entered into a second lottery to win a separate prize, also worth 3000 Rs. When participants were making their main hiring choices, they did not know that their neighbors would later be paid for predicting their answers. This rules out concerns that they tried to make their hiring choices more predictable in order to help out their neighbors.

G.3 Social desirability score

To measure the social desirability score of each participant in phase 1, I use an adapted version of the **?** index that includes the following questions:

1. I sometimes feel annoyed at people when I don't get what I want.
2. No matter who I'm talking to, I'm always a good listener (*reverse coded*).
3. I sometimes try to take revenge instead of forgiving and forgetting.
4. I am always polite, even to people who are not nice. (*reverse coded*)
5. There have been times when I was jealous of other people's luck.
6. I am sometimes annoyed when people ask me for favors.
7. I have deliberately said something that hurt someone's feelings.

This subset of questions was selected based on an exploratory factor analysis of pilot data. I calculate an individual's social desirability score by summing the number of socially desirable answers they give (that is, disagreeing with questions 1, 3, 5, 6, and 7, or agreeing with questions 2 and 4). This social desirability score is used in **??**.

G.4 Salience

I examine how salient the idea of transgender people is for each participant. I use a test of salience based on the one seen in **?**. Participants were read two lists containing a mix of words mostly related to deliveries, everyday objects, and identity. The first list contained the words: *Delivery, Dal, Tamil, Bucket, Sambar, Man, Water, App*, and *Insurance*. The second list contained

the words *Idly, Pot, Bike, Hindu, Hospital, Transgender, Butter, President*, and *Peas*. The lists were read out in the same order to every participant. After each list was read out once by the enumerator, participants were asked to repeat as many words as they could from the list. The enumerators were instructed to not repeat the options. To incentivize performance in the game, participants were truthfully told that if they recalled the most words of all the people in the study, they would be entered into a lottery with a prize worth Rs. 3000.

The measure of salience was whether they recalled the word "transgender", conditional on the total number of other words they recalled. In the *No discussion (private)* arm, people remembered the word "transgender" 75% of the time, and on average remembered other words 55% of the time. There is a significant correlation (p=0.04) between participants' recollection of the term "transgender" and their selection of a transgender individual in the outcome hiring round. This suggests that the salience measure is successfully capturing a signal that is relevant to hiring decisions.

G.5 List experiment

To measure negative attitudes towards transgender people, I use a double list experiment (**??**). In this method, participants are shown two lists of statements (list A and list B), and are asked how many statements from each list they agree with. They are not asked *which* statements they agree with, so neither the surveyor nor the researcher can determine whether they agreed with a particularly sensitive statement in the list. List A and B each contain 5 non-sensitive statements. For each participant, either list A or list B is randomly selected to include one additional statement: "In general, if I see a transgender person, I walk away." Enumerators read out each list and asked the participant how many statements they agreed with. Whether list A or list B was read first was also randomized. Using two lists has the advantage of enabling a validation check of the treatment effect estimates (**?**). Instead of pooling the treatment effect estimates across both lists, as in the main specification, I can estimate the treatment effect of the 3-person discussion separately for list A and list B. When using each list separately, the treatment effect estimates are similar (0.130 and 0.054 respectively), and the difference between the estimates is not significant (p=0.93).

G.6 Group relations

We asked participants questions about their relationships with others in their group, in order to understand how these affected group dynamics. In phase 1 of data collection, we asked each participant two questions about each of the other two people in their group: (i) What is your relationship with [*NAME*]? (ii) How well do you know [*NAME*]? (Options: *Very well*, *Quite well*, *Not very well*, *Very little*). I use question 1 to generate 4 dummy variables, indicating whether the other participant is (i) just a neighbor, (ii) a friend, (iii) a close family member, or (iv) another family member. In phase 2, I expanded this set to include the following additional questions: (iii) How long have you known [*NAME*]? (Options: *Less than 6 months, 6 months to 1 year, 1-5 years, 5+ years*); (iv) In general, how often do you talk to [*NAME*]? (Options: *Never, A few times per year, A few times per month, A few times per week, Most days, Every day*); (v) How often do you ask [*NAME*] for advice? (Same options as iv); (vi) How often do you ask [*NAME*] for recommendations of items to buy? (Same options as iv); (vii) How frequently do you tell secrets to [*NAME*]? (Same options as iv). I create an index of the perceived strength of the

relationship with another group participant. I use a factor analysis to generate loadings for the set of variables that includes the four dummy variables created by question 1, and the questions 2-7. I retain all measures that have a loading with an absolute value greater than 0.3. I create an index using a weighted sum of all measures where the weights are proportional to the estimated loadings. In cases where some data is missing (for example, phase 1 participants for whom we do not elicit questions 3-7), only the data that is present is used to calculate the weighted sum.

G.7 Private grocery pick up choices

Participants were told that they had been entered into a lucky draw to win a Rs. 5000 gift voucher, which could be used to buy grocery items. The winner would have to organize getting the items by calling the worker they selected, telling the worker which items they wanted, and meeting the worker at our office to pick up the items.[33] In this round, participants saw 4 pairs of options for who they could pick up the items from, and were told that if they won the lottery, we would randomly select one of their choices to organize the pickup with. 2 of the 4 pairs included a transgender worker.

The enumerator giving the interview did not know what responses were given. We did not ask the respondent for their choice verbally, as in the main hiring rounds. Instead, we gave the tablet directly to the respondent, and they clicked their preferred answer. Upon clicking, the tablet would automatically skip to the next question and not reveal again the answer chosen before, making it impossible for the enumerator to know what was selected. We truthfully told respondents that enumerators wouldn't know what was selected, making the answers anonymous.[34] The anonymity of their answers was well understood by the participants: only 0.9% said that their neighbors would know which options they picked, and only 1.1% said that the surveyor would know.

G.8 Persuasiveness

In phase 2 of data collection, we elicited a set of questions designed to measure how persuasive an individual was likely to be in a group discussion. For each question, the participant was asked to rate out of 10 how they scored on a measure of a personality trait. 5 of the traits measured are associated with extraversion and leadership, while 2 were associated with introversion. The questions were: (i) Out of 10, how confident is [NAME]?; (ii) Out of 10, how quiet is [NAME]? (reverse coded); (iii) Out of 10, how like a leader is [NAME]?; (iv) Out of 10, how shy is [NAME]? (reverse coded); (v) Out of 10, how talkative is [NAME]?; (vi) Out of 10, how admirable is [NAME]?; (vii) Out of 10, how inspiring is [NAME]?. These questions were selected from a broader set of questions by selecting the subset of questions that loaded onto the first factor in an exploratory factor analysis of pilot data. I combine the questions into a persuasiveness index by correcting for acquiescence bias, using a factor analysis with one factor to generate factor loadings for each of the 7 measures, and retain all measures to create an index using a weighted sum of all measures, where the weights are proportional

[33] In order to ensure that participants anticipated some extended face-to-face contact with the worker, they were also told that they had to have a 15-minute conversation with the worker to give feedback on the process.

[34] Although participants still presumably realized that their data could be used for research purposes, this elicitation nevertheless plausibly reduces the impact of social image concerns on their behavior because the salient social judge, the enumerator, would not know how they had answered.

to the estimated loadings. Each participant is rated by both their neighbors. The correlation between the two ratings for each person is positive and significant (Pearson's correlation of 0.18, $p<0.001$), even when controlling for rater fixed effects (Pearson's correlation of 0.16, $p<0.001$). This suggests that the rating detects a meaningful characteristic of the participant.

G.9 LASSO controls

Following **?**, I use double LASSO to select controls in the main results. The full set of possible controls that were selected from was:

(1) Female (=1); (2) Speaks English (=1); (3) Reads English (=1); (4) Hindu (=1); (5) Bachelor's degree (=1); (6) Married (=1); (7) Employed (=1); (8) Landlord (=1); (9) Num. children; (10) Employer (=1); (11) Household size; (12) Monthly household food expenditure per capita (Rs.); (13) Num. family members in group-of-3; (14) Num. neighbours in group-of-3; (15) Num. friends in group-of-3; (16) Taken part in market research survey (=1); (17) Has received free item as promotion (=1); (18) Someone in household ordered taxi with app (=1); (19) Someone in household ordered food with app (=1); (20) Someone in household ordered other items with app (=1); (21) Self-reported WTP for delivery; (22) Respondent would normally be household member who receives delivery (=1); (23) Relative number of items offered by worker; (24) Relative reliability score; (25) Reliability score is shown (=1); (26) Reliability score of the benchmark worker.

In addition, in interaction specifications where the main treatment was identified by the interaction *Worker is trans* $\times$ *Treatment*, I also include the controls interacted with *Worker is trans* as possible controls. I also calculate the mean of each control variable for the two other people in a participant's group-of-3, and include that mean as a possible control. When there are multiple treatment arms in one specification (e.g., for the phase 2 discussion-arm treatment arms), I include the union of the controls selected by a double LASSO using each of the treatment dummies. I indicate which controls were selected for Tables 1 and 2 by the LASSO selection process in **??**.

Table G1: *LASSO controls used in Table 1 and Table 2*

	Effect of 3-person discussion (Table 1)		Effect of rights videos (Table 2)	
Variable	(2)	(3)	(2)	(3)
Female (=1)	X		X	X
Group-level control: Age			X	
Group-level control: Bachelor's degree (=1)			X	X
Group-level control: Employed (=1)			X	
Group-level control: Employer (=1)	X	X		
Group-level control: Has received free item as promotion (=1)			X	
Group-level control: Landlord (=1)	X		X	
Group-level control: Married (=1)			X	
Group-level control: Monthly h.h. food expenditure per capita (Rs.)			X	
Group-level control: Num. children			X	
Group-level control: Relative number of items offered by worker	X			X
Group-level control: Reliability score of the benchmark worker			X	
Group-level control: Respondent would normally be household member who receives delivery (=1)			X	
Group-level control: Self-reported WTP for delivery	X		X	X
Group-level control: Someone in household ordered food with app (=1)	X		X	X
Group-level control: Someone in household ordered other items with app (=1)			X	
Group-level control: Taken part in market research survey (=1)	X			
Married (=1)	X		X	
Relative number of items offered by worker	X	X	X	X
Relative reliability score	X		X	
Worker is trans x Age			X	
Worker is trans x Hindu (=1)			X	
Worker is trans x Household size	X		X	
Worker is trans x Married (=1)			X	
Worker is trans x Monthly h.h. food expenditure per capita (Rs.)	X		X	
Worker is trans x Reads English (=1)			X	
Worker is trans x Self-reported WTP for delivery	X			
Worker is trans x Someone in household ordered other items with app (=1)			X	

Group-level control is the mean value of the variable for the other two people in a participant's group. (2) and (3) indicate the column numbers from Table 1 and Table 2 in the main text.

H Transcript data

H.1 Sentence-level analysis (k-means clustering)

I start with a hand-transcribed dataset where the observation is at the (discussion-choice × participant × sentence) level. This includes the full sentences spoken by the participant in Tamil and translated into English by research assistants. I use OpenAI's *embedding-3-small* model to extract a 1,536-dimensional semantic embedding vector for each English sentence. I then apply k-means clustering to the original high-dimensional embeddings to classify each sentence into one of 20 clusters. I set the number of clusters to 20, chosen to provide a granular view of different discussion topics while maintaining interpretable cluster sizes. For each cluster, I identify representative utterances by computing the cosine similarity between each utterance's embedding and its assigned cluster's centroid. Utterances with the highest similarity to their cluster centroid were selected as representative examples, shown in **??**, to provide interpretable characterizations of each cluster's semantic content.

H.2 Broader transcript features and AI hypothesis generation

To analyze the content of group discussions at the transcript level (where one transcript corresponds to the discussion about a given choice-pair), I used an iterative process using OpenAI's *GPT-4o-mini* large language model to generate and evaluate hypotheses about discussion patterns. The methodology consisted of three main stages: hypothesis generation, transcript rating, and dimensional reduction through factor analysis.

Hypothesis generation. I generated 500 hypotheses by randomly pairing two group discussion transcripts (for discussions involving a transgender worker) and asking *GPT-4o-mini* to identify the main difference between them. For each pair of randomly selected discussions, the following prompt was used:

```
BACKGROUND:
Below are transcripts from 2 group discussions in Chennai, India, in which 3 participants
    discussed whom they would prefer to hire to deliver groceries to their home. Participants had
    to choose between option A or B. They were shown photos of the two delivery workers, one of
    whom was male, and one of whom was transgender. Each discussion dealt with multiple rounds,
    changing the choice of Option A and Option B each time. The grocery items on offer were: Aachi
     masala spice, tea powder, and ghee. Option A and B may have offered the same set of items, or
     different sets of items.
They were also in some cases given information about:
 - the languages spoken by the delivery workers (only Tamil, or Tamil and English)
 - the delivery workers' experience
 - how many deliveries they completed in a training task
Participants were asked to discuss which option they preferred, and why, and then make a
    collective choice between the two options.

TRANSCRIPTS:
   The transcripts are in JSON format.
   Each discussion is represented as a list of utterances, with each utterance containing the
       following fields:
   - transcript_line_id: the order of the utterance in the discussion
   - who_speaking_label: the label of the participant speaking (1, 2, or 3, or unknown)
   - speech_english: the English translation of the participant's speech
   - pro_a_b: a column manually coded by research assistants indicates whether the participant is
       arguing to choose A or B, or neither

   TRANSCRIPT 1:
```

```
    {TRANSCRIPT 1}

  TRANSCRIPT 2:
    {TRANSCRIPT 2}

TASK:
Your task is to identify what has changed from TRANSCRIPT 1 to TRANSCRIPT 2. Focus on the
    generalizable insight that can be applied in other contexts. Ignore things that are specific
    to these transcripts. Do not make references to these transcripts that may not be relevant for
     others. Come up with an insight that captures the sort of change observed moving from
    TRANSCRIPT 1 to TRANSCRIPT 2.

Come up with an insight as a single sentence in this exact format:
Hypothesis: _____ is the main difference between TRANSCRIPT 2 compared to TRANSCRIPT 1.

Please make sure that the hypothesis is:
i. clear (i.e., precise, not wordy, and easy to understand);
ii. generalizable to novel situations (i.e., they would make sense if applied to other transcripts
    );
iii. empirically plausible (i.e., this is a dimension on which messages can vary);
iv. unidimensional (i.e., avoid hypotheses that list multiple constructs, so if there are many
    things changing, pick one);
v. usable (i.e., a human equipped with this insight could evaluate another group discussion in a
    similar way)
```

The temperature parameter for GPT-4 was randomized between 0.1 and 0.9 to ensure robustness to model parameters. Each generated hypothesis was then automatically cleaned to remove specific references to "Group 1" and "Group 2" in order to make them suitable to act as a rating. For example, "Hypothesis: The level of consensus among participants in decision-making is the main difference between Group 2 compared to Group 1." was converted to "The level of consensus among participants in decision-making."

Transcript rating. After generating the hypotheses, we used *GPT-4o-mini* to rate how well each of the 1034 discussion transcripts aligned with each of the 500 hypotheses on a scale of 1–10. Each transcript was rated exactly once against each hypothesis, using the following prompt:

```
BACKGROUND:
Below is a transcript from a group discussion in Chennai, India, in which 3 participants discussed
     who they would prefer to hire to deliver groceries to their home. Participants had to choose
    between option A or B.
They were shown photos of the two delivery workers, one of whom was male, and one of whom was
    either male, female, or transgender.
Each discussion dealt with multiple rounds, changing the choice of Option A and Option B each time
    .
The grocery items on offer were: Aachi masala spice, tea powder, and ghee. Option A and B may have
     offered the same set of items, or different sets of items.
They were also in some cases given information about:
 - the languages spoken by the delivery workers (Tamil only, Tamil and English)
 - the delivery workers' experience
 - how many deliveries they completed in a training task
Participants were asked to discuss which option they prefer, and why, and then make a collective
    choice between the two options.

TRANSCRIPT:
{TRANSCRIPT}

HYPOTHESIS:
```

```
{HYPOTHESIS}

TASK:
Rate how much this single transcript aligns with the hypothesis on a scale from 1 to 10, where:
  - 10 = maximally in line with the hypothesis
  - 1  = not at all in line with the hypothesis

Output in JSON with two keys:
  - explanation (string)
  - rating (integer, 1-10 or return NA for the rating if the transcript does not contain enough
      information to make a rating.)
```

As with hypothesis generation, the temperature parameter was randomized between 0.1 and 0.9 for each rating to ensure robustness to model parameters.

Dimension reduction. Since many of the 500 hypotheses generated were similar (e.g., "The level of consensus among participants in decision-making." and "The level of consensus among participants regarding their preference."), I group the hypotheses into highly correlated factors, based on the correlation of transcript ratings across hypotheses. I first used parallel analysis that suggested 11 factors in the data, corresponding to 11 hypothesis "types", and then use factor analysis to create indexes for these 11 hypothesis types. For each factor, I use the single hypothesis with the highest absolute loading as the label for that hypothesis type.

H.3 Morality ratings

Morality rating. I used OpenAI's *GPT-4o-mini* model to rate how much each discussion transcript invoked morality on a 0–1 scale. I repeated the rating 50 times with randomly-varying prompts to ensure robustness, and took the mean rating over all 50 ratings. This methodology, particularly the randomization of prompt elements and multiple analyses per transcript, was designed to minimize the impact of any particular prompt formulation or model parameter setting on the final results.

Prompt generation. The prompt was randomized along three dimensions:

1. Moral topics: Three topics were randomly selected from a pool of eleven possibilities: (i) *fairness* (ii) *rights* (iii) *justice* (iv) *giving people opportunities* (v) *ethical decision-making* (vi) *what is 'right' or 'wrong'* (vii) *responsibility and accountability* (viii) *personal autonomy* (ix) *care and compassion* (x) *human dignity* (xi) *truth and honesty*

2. Evaluation criteria: The inclusion of specific evaluation criteria was randomized, with two possible additions: (i) whether to explicitly consider both frequency and intensity of moral considerations; and (ii) whether to weight explicit moral references more heavily than implicit ones.

3. Temperature: The GPT-4o-mini temperature parameter was randomly set between 0.1 and 0.9 for each analysis to vary how "surprising" the model's outputs were.

Prompt template. The following base prompt was used, with randomized elements inserted as described above:

```
BACKGROUND:
Below is a transcript from a group discussion in Chennai, India, in which 3 participants (R1, R2,
    R3) discussed who they would prefer to hire to deliver groceries to their home. Participants
```

```
    had to choose between option A or B. They were shown photos of the two delivery workers, one
    of whom was male, and one of whom was either male, female, or transgender. Each discussion
    dealt with multiple rounds, changing the choice of Option A and Option B each time. The
    grocery items on offer were: Aachi masala spice, tea powder, and ghee. Option A and B may have
     offered the same set of items, or different sets of items.

They were also in some cases given information about:
- the languages spoken by the delivery workers (Tamil only, Tamil and English)
- the delivery workers' experience
- how many deliveries they completed in a training task

Participants were asked to discuss which option they prefer, and why, and then make a collective
    choice between the two options.

TASK: Analyze the transcript by rating how much the discussants invoke morality in their decision-
    making process on a scale from 0 to 1.

Rating Scale Details:
0: No moral considerations mentioned
1: Morality is the central theme and dominant focus

<randomly included for 50%>
The rating should reflect both:
- The frequency of mentions of moral considerations
- The intensity/emphasis of moral considerations when they appear
</randomly included for 50%>

<randomly included for 50%>
Explicit moral references (e.g., direct mentions of fairness, rights, justice, or 'right thing to
    do') should be weighted more heavily than implicit ones.
</randomly included for 50%>

Moral considerations include topics like:
- {RANDOM TOPIC 1}
- {RANDOM TOPIC 2}
- {RANDOM TOPIC 3}

Note: Simply mentioning demographic characteristics (like gender) without connecting them to moral
     concepts does not count as invoking morality.
```

I Heterogeneity in main results

Trade-off between items and worker. The pattern of choices indicates that participants traded off a preference for avoiding transgender workers with the value of the items on offer. Participants were sensitive to the items offered across each option in the pair: each additional item offered by one option in a pair made a participant 13 p.p. more likely to select that option (**??**, column 2). And people were less sensitive to items when shown a transgender person (**??**, column 1).[35]

Statistical discrimination. Belief-based (statistical) discrimination appears to underlie some of participants' unwillingness to select transgender people, driven by negative stereotypes that portray transgender workers as unreliable. Despite transgender workers having the same

[35]The sensitivity to items did not vary across treatment conditions (**??**, columns 2–4), which alleviates concerns that the collective nature of the choice made in the group discussion led to changes in preferences for bundles of goods that could confound the treatment effect on discrimination.

average reliability score as other genders in the experiment, participants rate transgender workers as less likely to complete a delivery (**??**, panel A, column 3; discussed below). To test whether this leads to discrimination, half of the choice-pairs included information about the reliability of both workers. Revealing the reliability score makes participants 2.9 p.p. more likely to select a transgender worker, and this effect is unique to transgender workers (**??**, column 1). Anti-transgender discrimination in the control group therefore appears to be partially driven by statistical discrimination. I cannot adjudicate whether it is "inaccurate" or "accurate" statistical discrimination (**?**) because participants may be accounting for features that are not accounted for in the experimental reliability score, such as the risk that a transgender worker is harassed and prevented from carrying out a delivery.

However, the effect of the discussion does not appear to be driven by changes in such statistical discrimination. The discussion does not significantly affect beliefs about the reliability of transgender workers (**??**, panel A, column 3). And I find no evidence that the 3-person discussion reduces the belief-based component of discrimination, although I am not well-powered for this test (**??**, column 2).[36]

Heterogeneity by participant gender. A heterogeneity analysis (**??**) shows that while anti-transgender discrimination is stronger for male participants than female participants (difference: 6.4 p.p., p=0.09), the *treatment effects* of the discussion are similar for both males and females (p=1.00). This is evidence against any explanations for the discussion's effects that are specific to a participant's gender. Relatedly, there is no significant treatment effect on preferences for cis-gender female delivery workers (estimate: 1.0 p.p., p=0.49). Nevertheless, the effect on male participants' preferences for female workers is substantial, even if insignificant (5.1 p.p., p=0.31), leaving open the possibility that the discussion might be reducing discrimination against female as well as transgender workers.

J Additional results on rights videos

Manipulation check. As a manipulation check, I show that participants' beliefs about the legal rights of transgender people (as measured by a summary index) are significantly affected by the legal rights video, but not by the rights messaging video (Appendix **??**).

Interaction effects. I do not find interaction effects between the *legal rights* video and group discussions (**??**); the reductions in discrimination caused by both combine approximately linearly ($p \in [0.96, 1.00]$). By contrast, there is weak evidence of a negative interaction effect between the *rights messaging* video and group discussions, such that the rights messaging video has no detectable effect on discrimination in the group-discussion arms ($p \in [0.66, 0.97]$). This may be because the content of the *rights messaging* video is very similar to the persuasive discourse in the discussion, therefore acting as a close substitute, whereas the *legal rights* video provides additional informational content.

Mechanism for legal rights videos. The effect of informing participants about the legal rights of transgender people appears to be partially mediated by changes in perceived norms and perceived reliability. The videos have a significant effect on perceived norms of discrimination

[36] While the point estimate of the interaction of (*Worker is trans* $\times$ *Reliability score is shown* $\times$ *3-person discussion*) is negative and large enough to negate the effect of (*Worker is trans* $\times$ *Reliability score*), I cannot reject that it is different from 0 (p=0.24).

Table I2: *Sensitivity to items does not vary across treatment arms and is lower for choices involving transgender workers*

	Chose worker in outcome round (=1)				Chose worker in treatment round (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Worker is trans × 3-person discussion	0.165***	0.164***	0.165***	0.165***	0.196***	0.196***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.030)	(0.030)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
3-person discussion	−0.001	−0.001	−0.001	−0.001	0.005	0.005
	(0.010)	(0.010)	(0.010)	(0.010)	(0.021)	(0.021)
	[0.938]	[0.944]	[0.913]	[0.916]	[0.800]	[0.800]
Relative # items offered	0.144***	0.138***	0.124***		0.132***	
	(0.006)	(0.008)	(0.007)		(0.008)	
	[<0.001]	[<0.001]	[<0.001]		[<0.001]	
3-person discussion × Relative # items offered		0.013	0.011		−0.026	
		(0.013)	(0.011)		(0.018)	
		[0.298]	[0.326]		[0.143]	
Relative # items offered × Worker is trans	−0.046***	−0.042***				
	(0.010)	(0.013)				
	[<0.001]	[<0.001]				
3-person discussion × Relative # items offered × Worker is trans		−0.009				
		(0.019)				
		[0.629]				
Relative value of items offered (Rs. / 100)				0.146***		0.153***
				(0.008)		(0.009)
				[<0.001]		[<0.001]
3-person discussion × Relative value of items offered (Rs. / 100)				0.012		−0.030
				(0.013)		(0.021)
				[0.334]		[0.143]
Num. observations	13 494	13 494	13 494	13 494	8996	8996
Num. participants	2249	2249	2249	2249	2249	2249
Num. groups	751	751	751	751	751	751
Controls	X	X	X	X	X	X
Controls interacted with worker is trans	X	X	X	X	X	X

Notes: *Relative # of items offered* is the number of items (1, 2 or 3) offered by the alternative worker, less the number of items offered by the male benchmark worker. *Relative value of items offered* is the relative cost in rupees of the items offered by the alternative worker compared to the benchmark worker, divided by 100 (to ease interpretation).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes *No discussion (private)* arm and *3-person discussion arm* in both phase 1 and phase 2 of data collection. In all columns the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was selected. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. Columns (1)-(4) show the private choices in the *outcome round*. Columns (5) and (6) show choices in the treatment round (for those in the discussion arm, this was the choices made *during* the discussion. The specification used is seen in equation **??**. Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; the relative reliability score; a dummy for whether the reliability score was shown; phase fixed effects; and the controls selected by double LASSO (see Section **??**). Controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

Table I3: *Effect of discussion on norms, attitudes, and beliefs about reliability*

Panel A: Norms

	Predicted share of people that pick trans (community)	Predicts that other picks trans (=1) (within group-of-3)
	(1)	(2)
3-person discussion	0.043***	0.240***
	(0.012) [<0.001]	(0.022) [<0.001]
Num. observations	2249	4465
Num. participants	2249	2238
Num. groups	751	751
Outcome mean: No discuss. (priv.)	0.50	0.36
Controls	X	X
q-value of treatment effect	0.001	0.001

Panel B: Attitudes and beliefs about reliability

	# statements agreed with (list experiment)	Disapproves of discrimination (=1)	Likely or very likely to complete delivery (=1)
	(1)	(2)	(3)
Anti-trans statement in list × 3-person discussion	0.071		
	(0.055) [0.216]		
Anti-trans statement in list	0.204***		
	(0.033) [<0.001]		
3-person discussion		0.017**	
		(0.008) [0.046]	
Photo is trans × 3-person discussion			0.035
			(0.026) [0.148]
Photo is trans			−0.086***
			(0.025) [0.001]
Num. observations	4498	4498	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Outcome mean: No discuss. (priv.)	2.90	0.93	0.71
Question FEs	X	X	X
Participant FEs	X		X
Controls	X	X	X
q-value of treatment effect	0.107	0.088	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes only the *No discussion (private)* and *3-person discussion* arms, in both phases. Controls include stratum fixed effects; dummies for the rights-video treatments; phase fixed effects; and the controls selected by double LASSO (see Section **??**). For Panel B, column (2), I include controls for the difference in items offered, the relative reliability score, and whether the reliability score is shown. As pre-specified, columns (1) and (2) are adjusted for multiple hypothesis testing using the q-value that controls for the false discovery rate (**?**).

Panel A, Column (1): Outcome is the incentivized predicted proportion of other people (out of 20) in the study will pick a transgender worker. Only the choice involving the transgender worker is included.

Panel A, Column (2): The unit of observation is the participant × prediction. Outcome is whether the participant predicted that another person in their group selected a transgender worker in the private outcome round. Only predictions involving a transgender worker are included.

Panel B, Column (1): Outcome is the number of statements the participant agreed with on a list of statements. Each participant sees both List A and List B. The anti-trans statement ("In general, if I see a transgender person, I walk away") is randomly included in either List A or List B. *Question FEs* is a fixed effect for List B.

Panel B, Column (2): Enumerator describes two discriminatory scenarios. Outcome is whether the participant says the person's actions are wrong. *Question FEs* is a fixed effect for the second scenario.

Panel B, Column (3): Outcome is whether the participant says a worker is likely or very likely to complete a delivery after being shown a photo. Participants rate two workers, one of whom is transgender. Order is randomized. *Question FEs* controls for the order of the choice.

Table I4: *Evidence of statistical discrimination against transgender workers*

	Chose worker in private outcome round (=1)	
	(1)	(2)
Worker is trans × 3-person discussion	0.173***	0.192***
	(0.022) [<0.001]	(0.027) [<0.001]
Worker is trans	−0.200***	−0.209***
	(0.038) [<0.001]	(0.039) [<0.001]
3-person discussion	0.000	−0.010
	(0.010) [0.967]	(0.014) [0.475]
Relative reliability score	0.020***	0.016***
	(0.004) [<0.001]	(0.005) [<0.001]
Reliability score is shown (=1)	0.012	0.004
	(0.010) [0.214]	(0.012) [0.735]
Worker is trans × Relative reliability score	−0.007	−0.012
	(0.007) [0.308]	(0.008) [0.143]
Worker is trans × Reliability score is shown (=1)	0.029*	0.043**
	(0.015) [0.052]	(0.020) [0.033]
3-person discussion × Relative reliability score		0.009
		(0.008) [0.237]
3-person discussion × Reliability score is shown (=1)		0.020
		(0.020) [0.318]
Worker is trans × 3-person discussion × Relative reliability score		0.014
		(0.013) [0.299]
Worker is trans × 3-person discussion × Reliability score is shown (=1)		−0.035
		(0.030) [0.242]
Num. observations	13 494	13 494
Num. participants	2249	2249
Num. groups	751	751
Controls	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was selected in the private choices in the *outcome round*. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. Controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; the relative # items offered; and the controls selected by double LASSO (see Section **??**). Relative reliability score is the reliability score (out of 10) of the alternative worker minus the benchmark worker. *Reliability score is shown* is 1 when the reliability score is shown. Relative reliability score is coded as 0 when it is not shown.

Table I5: *Heterogeneity by demographic characteristics*

	Chose trans in outcome round (=1) (pairs with trans only)	
	Uninteracted term	Interacted term (x 3-person discussion)
	(1)	(2)
Age	0.003***	−0.004*
	(0.001)	(0.002)
Female (=1)	0.064*	0.000
	(0.038)	(0.064)
Speaks English (=1)	−0.019	0.040
	(0.040)	(0.067)
Reads English (=1)	−0.004	0.010
	(0.033)	(0.052)
Hindu (=1)	0.076**	−0.063
	(0.029)	(0.049)
Bachelor's degree (=1)	0.005	−0.032
	(0.029)	(0.051)
Married (=1)	0.026	−0.032
	(0.033)	(0.055)
Employed (=1)	0.053*	−0.094*
	(0.031)	(0.050)
Landlord (=1)	−0.024	0.095
	(0.038)	(0.059)
Has children (=1)	0.025	−0.039
	(0.025)	(0.041)
Employer (=1)	−0.017	0.079*
	(0.025)	(0.042)
Above med. hh size (=1)	0.041*	−0.044
	(0.024)	(0.039)
Above med. hh food exp. p.c. (=1)	−0.007	0.016
	(0.022)	(0.038)
3-person discussion	0.402***	
	(0.107)	
Num. observations	4452	4452
Num. participants	2249	2249
Num. groups	751	751
Controls	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The columns together show the results from one regression. Column 1 shows the coefficients without interaction with *3-person discussion*. Column 2 shows the coefficients when interacted with *3-person discussion*. The outcome is whether the transgender worker was selected in the private outcome round, restricting analysis to only choices that include a transgender worker. Additional controls include stratum fixed effects; dummies for the rights videos; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; phase fixed effects; relative # items offered; relative reliability score; whether the reliability score was shown.

(**??**, columns 1–2). Participants predict that others will select transgender workers more, both in the wider community (2–3 p.p.) and in their group of 3 (4–6 p.p.). The videos also lead to small increases in whether a transgender worker is deemed likely to complete the delivery (6 p.p., 8%, **??**, column 5). By contrast, neither video has a detectable effect on attitudes, as measured by the list experiment or the questions on disapproval of discrimination (**??**, columns 3–4). The rights videos' effects could also be driven by fear of anti-discrimination laws being enforced against participants.

Table J6: *Legal rights video affects beliefs about the legal status of transgender people*

	Say trans have legal status (=1)	Say trans have legal status + correctly name at least one legal right (=1)	Number of legal rights correctly named	Not employing is illegal (=1)	Avoiding on street is illegal (=1)	Summary index (Z)
	(1)	(2)	(3)	(4)	(5)	(6)
Rights messaging video	0.009	0.038*	0.200***	−0.004	−0.013	0.034
	(1.4×10^{-2}) [0.525]	(2.0×10^{-2}) [0.055]	(5.0×10^{-2}) [<0.001]	(1.5×10^{-2}) [0.802]	(1.8×10^{-2}) [0.467]	(2.8×10^{-2}) [0.218]
Legal rights video	0.098***	0.195***	0.890***	0.034**	0.034**	0.269***
	(1.1×10^{-2}) [<0.001]	(1.8×10^{-2}) [<0.001]	(5.4×10^{-2}) [<0.001]	(1.4×10^{-2}) [0.016]	(1.7×10^{-2}) [0.044]	(2.6×10^{-2}) [<0.001]
Num. participants	3397	3397	3397	3397	3397	3397
Num. groups	1134	1134	1134	1134	1134	1134
Outcome mean: Control vid.	0.87	0.64	1.11	0.85	0.79	0.00
Controls	X	X	X	X	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Unit of observation is the participant. Sample includes all participants in all discussion-arm treatments, in both phase 1 and 2 of data collection. Controls include stratum fixed effects; dummies for the discussion-arm treatments; whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; phase fixed effects; and the controls selected by double LASSO (see Section **??**). *Say trans have legal status* is an indicator for whether the participant responds yes to "Do transgender people have legal status?". *Correctly name at least one legal right* indicates whether the participant was able to correctly name one legal right that transgender people hold in India in response to the question "What legal status do transgender people have?". *Number of legal rights correctly named* is the number of correct legal rights named in response to this same question (coded as 0 if they say that transgender people do not have legal status). *Not employing is illegal*: after listening to a discriminatory vignette ("Two people approach someone for a job: a man and a transgender. The employer rejects the transgender because they are transgender."), the participant said that the employer is breaking the law. *Avoiding on street is illegal*: after listening to a second discriminatory vignette ("A woman avoids a transgender person on the street, because they are transgender."), the participant said that the woman is breaking the law. *Summary index (Z)* is created by (i) normalizing each of the outcome variables in columns 1, 3, 4, and 5 by subtracting from the control-video mean and dividing by the control-video standard deviation; (ii) combining these normalized variables into an index with weights based on the inverse-covariance matrix (**?**).

Table J7: *Interactions between trans rights videos and discussions*

	Chose trans in private outcome round (pairs with trans only) (=1)	
	3-person discussion + No discussion (private) (Phases 1 + 2)	All discussion arms except listeners (Phase 2 only)
	(1)	(2)
Rights messaging video	0.070**	0.110***
	(0.028) [0.012]	(0.037) [0.003]
Legal rights video	0.060**	0.118***
	(0.027) [0.027]	(0.037) [0.001]
3-person discussion	0.193***	0.231***
	(0.035) [<0.001]	(0.057) [<0.001]
Rights messaging video × 3-person discussion	−0.081	−0.144*
	(0.050) [0.109]	(0.080) [0.070]
Legal rights video × 3-person discussion	0.003	0.000
	(0.048) [0.957]	(0.075) [0.996]
No discussion (public)		0.032
		(0.039) [0.422]
2-person discussion (listener)		0.189***
		(0.056) [<0.001]
2-person discussion (speaker)		0.154***
		(0.050) [0.002]
Rights messaging video × No discussion (public)		0.002
		(0.058) [0.971]
Rights messaging video × 2-person discussion (listener)		−0.110
		(0.079) [0.162]
Rights messaging video × 2-person discussion (speaker)		−0.083
		(0.070) [0.238]
Legal rights video × No discussion (public)		0.017
		(0.058) [0.773]
Legal rights video × 2-person discussion (listener)		−0.075
		(0.082) [0.357]
Legal rights video × 2-person discussion (speaker)		0.023
		(0.067) [0.727]
Num. observations	4498	4436
Num. participants	2249	2218
Num. groups	751	741
Controls	X	X
p-val: (Rights messaging video | 3-person discussion)	0.822	0.656
p-val: (Rights messaging video | No discussion (public))		0.012
p-val: (Rights messaging video | 2-person discussion (listener))		0.967
p-val: (Rights messaging video | 2-person discussion (speaker))		0.635
p-val: (Legal rights video | 3-person discussion)	0.100	0.071
p-val: (Legal rights video | No discussion (public))		0.003
p-val: (Legal rights video | 2-person discussion (listener))		0.540
p-val: (Legal rights video | 2-person discussion (speaker))		0.012

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Outcome is whether a participant chose the transgender worker in the private outcome round (restricting analysis to only choices with transgender workers). Sample in columns (1) and (2) includes only the *3-person discussion* arm and the *No discussion (private)* arm, in both phases. *p-val: (Rights messaging video | 3-person discussion)* denotes the *p*-value on the test that the effect of the rights messaging video is 0 for participants in the 3-person discussion arm. Other p-values are defined analogously. Controls include stratum fixed effects; phase fixed effects (columns 1 and 2 only); whether the individual was randomized into being offered 3 deliveries or 1 delivery, or was not part of this randomization; whether the alternative worker was shown on the right; and the controls selected by double LASSO (see Section **??**).

K Further evidence on persuasion

Amount of conversation. If participants updated their personal normative beliefs or their beliefs about the prescriptive norm due to the discussion, we would expect *more conversation* to lead to less discrimination (either because there are more signals about the norm, or because participants are exposed to more persuasive messages). The correlational evidence supports this: a 1 standard deviation increase in how much conversation there was about transgender

Table J8: *Effect of rights video on mechanism outcomes*

	Predicted % who pick trans (community)	Predicted % who pick trans (within group-of-3)	# statements agreed with (list experiment)	Disapproves of discrimination (=1)	Likely or very likely to complete delivery (=1)
	(1)	(2)	(3)	(4)	(5)
Rights messaging video	0.023**	0.045**		−0.007	
	(0.011) [0.045]	(0.021) [0.033]		(0.009) [0.417]	
Legal rights video	0.027**	0.066***		0.011	
	(0.011) [0.015]	(0.020) [0.001]		(0.008) [0.173]	
Anti-trans statement in list			0.185***		
			(0.050) [<0.001]		
Anti-trans statement in list × Rights messaging video			−0.067		
			(0.053) [0.212]		
Anti-trans statement in list × Legal rights video			0.010		
			(0.052) [0.846]		
Photo is trans					−0.099***
					(0.030) [<0.001]
Photo is trans × 3-person discussion					0.024
					(0.021) [0.259]
Photo is trans × Rights messaging video					0.058**
					(0.025) [0.022]
Photo is trans × Legal rights video					0.055**
					(0.025) [0.030]
Num. observations	3397	6741	6794	6794	6794
Num. participants	3397	3377	3397	3397	3397
Num. groups	1134	1133	1134	1134	1134

Notes: * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Sample includes all participants in both phases.
Column (1): The unit of observation is the participant. The dependent variable is the incentivized prediction of the proportion of other people (how many out of 20) in the study who pick a transgender person to receive a delivery when shown a specific pair of workers. Each participant makes 3 incentivized predictions, one of which includes a transgender worker. Only the choice involving the transgender worker is included for analysis.
Column (2): The unit of observation is the participant × prediction. The dependent variable is whether the participant predicted that another person in their group selected a transgender worker in the private outcome round. The prediction is incentivized. Each participant made 2 predictions (one involving a transgender worker) for each of their 2 group members. The two predictions involving a transgender worker are included for analysis. Controls include stratum fixed effects; dummies for the discussion-arm treatments; phase fixed effects; and the controls selected by double LASSO (see Section **??**).

workers (as rated by the enumerator) was correlated with listeners being 12 p.p. more likely to select a transgender worker (p=0.002).

Disagreement. In line with the claim that the arguments made about transgender workers are more persuasive than those made about non-transgender workers, groups are less likely to *disagree* when discussing transgender workers. Research assistants coded whether, during a discussion, some participants disagreed by arguing for opposing options. For choices involving a transgender worker, only 16% disagreed (compared to 21% for non-transgender pairs, p of difference=0.04), possibly indicating more persuasive arguments.

L Alternative mechanisms

In this section, I document evidence against a number of other mechanisms that might underlie the treatment effect of the discussion and the rights videos.

Other photo characteristics. If the transgender worker photos were observably different from non-transgender photos, this could have driven some of the treatment effects. For example, if transgender workers looked *poorer*, the discussion's effect might be driven by changing preferences for hiring low-income workers. To evaluate this concern, I used a separate sample of 500 online respondents from Tamil Nadu to rate the characteristics of a set of 30 photos used in the study. Participants rated photos in terms of perceived income, religion, age, caste, education level, and how neatly workers were dressed. They also rated how comfortable they would be talking to the worker, how unsafe they would feel having the worker in their home, how worried they would be if the worker spoke to their family, and how unhappy their spouse would be if the participant spoke to the worker. There were substantial differences in the perceived characteristics of transgender workers compared to non-transgender workers — e.g., 28% of transgender photos were rated as being very likely to come from a Scheduled Caste, compared to only 19-20% for male and female photos. Participants also perceive transgender workers as relatively poorer, more likely to be Hindu, less educated, and less neatly dressed compared to other workers. However, after controlling for the perceived characteristics of the worker photo, the results do not change qualitatively: the discussion still reduces discrimination by an estimated 20 p.p. (**??**). This suggests that the treatment effects are driven by changes in preferences for selecting transgender workers *per se*, rather than by changes in preferences for any correlated characteristics such as caste or age.

Social image concerns that continue in the outcome round. Even when participants made hiring choices in private in the outcome round (without their neighbors listening), their choices may have been affected by social image concerns. Knowing that their neighbors might see who delivered groceries to their home, they might choose a transgender worker to signal that they were non-discriminatory to their neighbors. To assess whether the treatment effects remained when shutting down this channel, I use a series of supplementary hiring choices. These *private grocery pick-up* choices (detailed in Appendix **??**) were designed to be more robustly private than the main outcome in two ways. First, participants had to pick up grocery items from the team office instead of receiving the delivery at home, so that neighbors would not know which worker was chosen. Second, I adjusted the elicitation process so that the participants' responses were hidden from the surveyor giving the interview.

The 3-person discussion still reduced discrimination for the private grocery pick-up choices (**??**).[37] The discussion treatment effect on this outcome is large, although slightly smaller in magnitude than the main hiring outcome (11.7 p.p., p=0.003). The *legal rights* video also reduces discrimination significantly, with a similar magnitude to the main outcome (9 p.p., p<0.001). Taken together, the results suggest that social image concerns *after* the discussion has ended are not sufficient to explain the measured treatment effects, although I cannot rule

[37] Discrimination in the *No discussion (private)* arm was stronger for these private outcomes than for the main hiring elicitation (29.1 p.p., p<0.001). The more extreme discrimination may come from a perception of increased intensity of social contact between the participant and the chosen worker: the participant was told they would have to speak on the phone to the worker and then organize a time to come to the office *alone* and speak to them for 15 minutes.

out that such concerns play some role.

Contemplation. Discussions may change people's hiring choices by making them think more carefully about their choices, or by allowing them to override an automatic discriminatory response (**????**). There is some evidence for such increased contemplation. Discussion participants take on average 2.2 seconds (27%) longer in the individual outcome-round choices, and are *less* likely to select a dominated option in the outcome round if they have been in a group discussion (p=0.02, **??**, column 1), suggesting they are being more attentive. However, it is unlikely that this drives the treatment effects on discrimination, since longer response times are not correlated with being more likely to select a transgender worker (p=0.43).

Experimenter demand effects. If participants wanted to please the surveyors or researchers, then those who correctly guessed the purpose of the study may have discriminated less against transgender workers (**?**). To measure this, we asked respondents to report their beliefs about the purpose of the study twice during the main survey (immediately after the hiring choices, and again at the very end of the session) using an open-ended question. I classify people as having correctly guessed the study's purpose if they mentioned transgender people. I find no evidence that experimenter demand effects confound the main treatment effects. 8% of participants correctly guess the purpose of the study after the main hiring round, and 12% correctly guess it by the end of the survey. However, discussion participants are no more likely to guess the purpose of the study at either stage than the control participants (**??**, columns 2 and 3), and in fact are more likely to think that the study is about delivery preferences (**??**).[38] Moreover, there is no detectable difference in the treatment effects for people who do and do not correctly guess the study purpose (**??**, columns 1 and 2).

While the rights videos did increase the likelihood of a participant correctly guessing the purpose of the experiment from 5% to 10% ($p<0.001$), those who correctly guessed did not drive the reductions in discrimination seen in the discussion groups (**??**, columns 1 and 2). These tests do not fully rule out *subconscious* demand effects, but the *Legal rights video* likely represents the upper bound on such demand effects, and has a substantially smaller treatment effect than the *3-person discussion*, suggesting that the discussion's effects are not driven by experimenter demand.

Social desirability bias. To measure a participant's propensity to give socially desirable answers, at baseline I elicited a shortened version of the **?** module, which has been used elsewhere in India for a similar purpose (**?**). The questions ask whether the respondent has a number of "too good to be true" traits (see Appendix **??**). I find no evidence that the results are driven by a participant's desire to give socially desirable answers to the enumerator. The treatment effects of the discussion and the rights videos are not significantly larger for individuals with an above-median social desirability score (**??**, column 3; **??**, column 3).

Salience. Simply increasing the salience of the idea of being transgender does not appear to be the key driver of the treatment effects. To measure this, I included a recall task in which participants have to restate as many items as possible from two lists of items, one of which includes the word "transgender". The probability of recalling the word transgender,

[38]This does not mean participants were unaware that they were discussing transgender workers with their group members, only that they did not think that this was the purpose of the study.

conditional on the number of other items recalled, is used to measure the salience of the idea of being transgender. Salience actually decreases in the 3-person discussion arm (**??**, column 1), and the effect on discrimination is not significantly stronger for participants who remembered the word transgender (**??**, column 4).

Cheap talk. To examine the robustness of the results to variation in the stakes, for a subsample of 582 individuals in phase 1 of data collection, I cross-randomized whether the participants were (truthfully) told that they would receive 1 delivery (N=288) or 3 deliveries (N=294) from the *same* worker. If the results were driven by experimenter demand effects, or by social image benefits that outweigh the cost of a *single* interaction with a trans worker, then receiving 3 deliveries would reduce the treatment effect of the discussion. While the people who are offered 3 deliveries discriminate more on average, the reduction in discrimination due to the discussion is still large and robust in the 3-delivery case (**??**, 14 p.p., p=0.013), and the interaction between the treatment effect and the number of deliveries is close to 0 and insignificant (p=0.79). The main effects of the discussion are therefore unlikely to be driven by the relatively low stakes of a single interaction.

Facilitator influence. One concern about the experimental design is that the way the communication was structured—mediated through a facilitator—may have inhibited anti-trans expression and driven the positive treatment effects. While it is challenging to test this directly (the enumerator who facilitates the discussion is always present), the pattern of heterogeneity does not support this interpretation. Using the number of words said by facilitators in the discussion as a proxy of how much influence they exert, I find that when facilitators influence the discussions about transgender workers more, this is associated with participants being *less* likely to choose a transgender worker (**??**). This correlational pattern appears incompatible with facilitator influence driving the positive effects. Instead, the reduced form estimate may be a lower bound of the true effect if facilitators endogenously aim to influence groups more in cases where the group appears to be initially more discriminatory.

Table L9: *Discussion effect is robust to controlling for other photo characteristics*

	Chose worker in private outcome round (=1)	
	(1)	(2)
Worker is trans × 3-person discussion	0.200***	0.200***
	(0.035) [<0.001]	(0.035) [<0.001]
Worker is trans	−0.193***	−0.193***
	(0.025) [<0.001]	(0.025) [<0.001]
3-person discussion	−0.017	−0.017
	(0.018) [0.363]	(0.018) [0.349]
Diff. in perc. wealth (Z)	0.000	−0.027
	(0.018) [0.989]	(0.023) [0.256]
Diff. in perc. age (Z)	−0.021	−0.024
	(0.013) [0.118]	(0.018) [0.169]
Diff. in perc. Scheduled Caste (Z)	−0.011	−0.028*
	(0.013) [0.423]	(0.017) [0.094]
Diff. in perc. educated (Z)	−0.011	−0.005
	(0.025) [0.656]	(0.033) [0.877]
Diff. in perc. neatly dressed (Z)	0.011	0.021
	(0.017) [0.496]	(0.021) [0.309]
Diff. in comfort talking (Z)	−0.005	0.011
	(0.026) [0.836]	(0.035) [0.746]
Diff. in feeling unsafe at home (Z)	−0.044*	−0.011
	(0.023) [0.061]	(0.029) [0.693]
Diff. in worried about talking to family (Z)	0.015	−0.014
	(0.023) [0.515]	(0.032) [0.647]
Diff. in spouse unhappy if talking (Z)	0.034**	0.048***
	(0.014) [0.016]	(0.018) [0.010]
3-person discussion × Diff. in perc. wealth (Z)		0.062*
		(0.037) [0.089]
3-person discussion × Diff. in perc. age (Z)		0.011
		(0.026) [0.676]
3-person discussion × Diff. in perc. Scheduled Caste (Z)		0.043
		(0.027) [0.108]
3-person discussion × Diff. in perc. educated (Z)		−0.009
		(0.051) [0.865]
3-person discussion × Diff. in perc. neatly dressed (Z)		−0.026
		(0.032) [0.421]
3-person discussion × Diff. in comfort talking (Z)		−0.043
		(0.053) [0.422]
3-person discussion × Diff. in feeling unsafe at home (Z)		−0.078*
		(0.044) [0.081]
3-person discussion × Diff. in worried about talking to family (Z)		0.069
		(0.047) [0.143]
3-person discussion × Diff. in spouse unhappy if talking (Z)		−0.033
		(0.027) [0.209]
Num. observations	4213	4213
Num. participants	2249	2249
Num. groups	751	751

Notes: This table shows the effect of the discussion on the probability of choosing the alternative worker when controlling for the *characteristics* of the photos. Photo characteristics were measured using a supplementary online survey (Dec 2023–Jan 2024), in which a sample of 500 new participants reported their perceptions of whether worker photos looked like they were rich, old, from a scheduled caste/tribe, educated, their most likely religion, and whether they were neatly dressed, all using 4-option Likert scales. They also rated photos based on whether they would (i) feel comfortable talking to the worker; (ii) feel unsafe if the worker visited their home; (iii) feel worried if the worker spoke to their family; (iv) think that their spouse would be unhappy if they spoke to the worker. Participants were recruited using Facebook advertisements, were 50% female, and were all current residents of Tamil Nadu. A subset of 30 photos (10 male, 10 female, 10 transgender) were rated. Each photo received between 74 and 98 ratings. Ratings were converted into Z-scores. The explanatory variables used are the *differences* in the Z-scores between the alternative worker and the benchmark worker. The outcome is whether the participant selected the alternative worker in the private outcome round. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Sample is phase 1 and 2, only No discussion (private) and 3-person discussion arms.

Table L10: *Effect of discussion on private grocery pick-up choices (phase 2 only)*

	Chose worker in private pick-up round (=1)		Chose trans in private pick-up round (=1) (trans pairs only)
	(1)	(2)	(3)
Worker is trans	−0.291***		
	(0.020) [<0.001]		
Worker is trans × 3-person discussion	0.117***	0.115***	
	(0.039) [0.003]	(0.038) [0.002]	
3-person discussion	0.011	0.010	0.123***
	(0.023) [0.639]	(0.023) [0.650]	(0.030) [<0.001]
Worker is trans × Speaker (2-person discussion)	0.109***	0.109***	
	(0.036) [0.003]	(0.036) [0.002]	
Speaker (2-person discussion)	0.021	0.018	0.126***
	(0.022) [0.344]	(0.021) [0.392]	(0.029) [<0.001]
Worker is trans × Listener (2-person discussion)	0.113***	0.113***	
	(0.042) [0.007]	(0.041) [0.006]	
Listener (2-person discussion)	0.022	0.016	0.125***
	(0.027) [0.415]	(0.026) [0.532]	(0.033) [<0.001]
Worker is trans × No discussion (public)	0.032	0.031	
	(0.029) [0.269]	(0.029) [0.283]	
No discussion (public)	−0.021	−0.020	0.009
	(0.019) [0.273]	(0.018) [0.280]	(0.022) [0.678]
Num. observations	8872	8872	4436
Num. participants	2218	2218	2218
Num. groups	741	741	741
Outcome mean†: no discuss (priv.), worker non-trans	0.63	0.63	
Outcome mean†: no discuss (priv.), worker trans	0.34	0.34	0.34
Controls		X	X
Controls interacted with worker is trans		X	

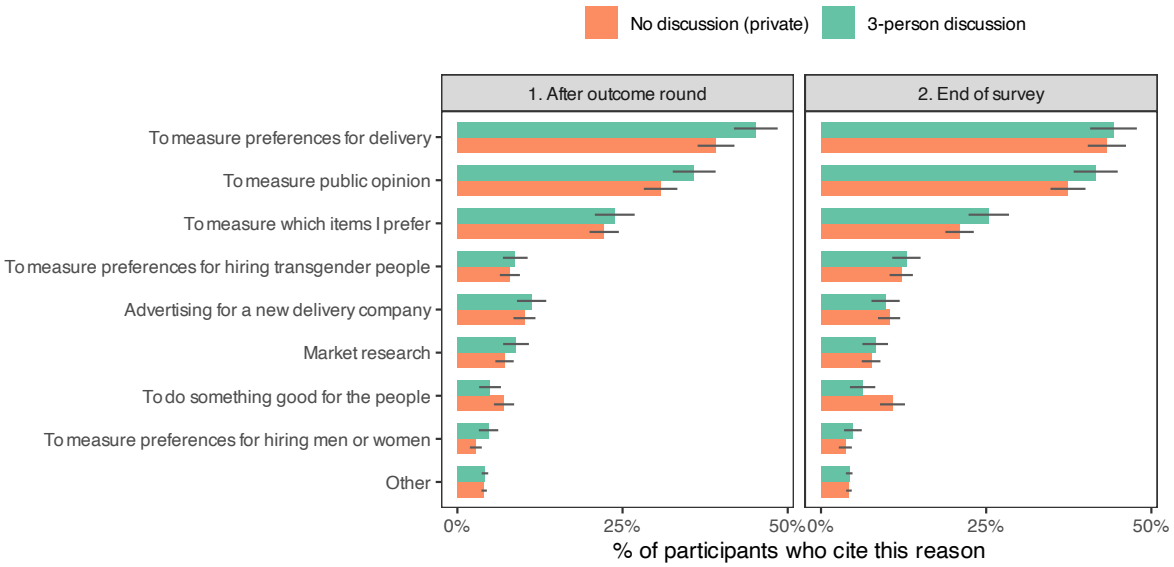
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all participants in phase 2. Column (3) only includes choices that involved a transgender worker. Participants saw 4 options, and were asked which worker they would prefer to organize a private grocery pick-up with. Neither the enumerator nor a participants' group members knew what they selected. In columns (1) and (2), the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) was selected the private grocery pick-up round. In column (3), it is whether the transgender worker was selected. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. The specification used is seen in equation **??**. Controls include stratum fixed effects; dummies for the discussion-arm treatments; whether the alternative worker was shown on the right; phase fixed effects; and the controls selected by double LASSO (see Section **??**). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

Table L11: *Discussion participants are not more likely to guess purpose of the experiment and are less likely to remember the word "transgender"*

	Remembered word 'transgender' (=1) (Phase 1 only)	Correctly guess purpose (=1) (after main outcome)	Correctly guess purpose (=1) (end of experiment)
	(1)	(2)	(3)
3-person discussion	−0.041*	0.007	0.007
	(0.024) [0.097]	(0.012) [0.542]	(0.013) [0.620]
Proportion of non-trans words remembered	0.189**		
	(0.080) [0.019]		
Num.Obs.	1179	2249	2249
R2	0.014	0.033	0.052
R2 Within	0.006	0.014	0.023
R2 Within Adj.	0.004	0.013	0.021
N participants	1179	2249	2249
N groups	393	751	751
LASSO controls	X	X	X
Mean: No discussion (private)	0.747927031509121	0.0794701986754967	0.121412803532009
Mean: 3-person discussion	0.706597222222222	0.0876404494382022	0.129213483146067
FE: phase		X	X
Controls	X	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets, and use randomization inference for the *3-person discussion* coefficients. Unit of observation is the participant level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm. Column (1) includes only phase 1, since salience module was only included in phase 1. Columns (2) and (3) include both phases 1 and 2. *Column (1).* Participants were read two lists of words, described in Section **??**, and were asked to recall as many of the words as possible. Outcome is whether the participant remembered the word transgender. I control for the proportion of other words remembered. *Columns (2) and (3).* Participants were asked what they thought the purpose of the study was twice: once after the main outcome round (column 2), and again at the end of the session (column 3). I classify people as having correctly guessed the study's purpose if they say it is to measure preferences for hiring transgender individuals. Outcome is whether they correctly guessed the purpose of the study.
Additional controls include: stratum fixed effects; phase fixed effects (for columns 2 and 3 only); dummies for rights videos; and controls selected by double LASSO (see Section **??**).

Figure L12: *Perceived purpose of the experiment*



Notes: Unit of observation is the participant level. Participants are asked what they believe the purpose of the study is twice: once immediately after the main hiring outcome round, and again at the end of the survey. Outcome on the y-axis is whether the participant cited the reason. Confidence intervals are based standard errors that are clustered at the group-of-3 level. To test whether the composition of perceived purposes changes, I regress the treatment status on indicator variables for each of the perceived purposes. The joint F stat for the coefficient on all the indicator variables is 2.5 (p=0.002) for after the outcome round, and 1.6 (p=0.09) for the end of the survey.

Table L13: *Treatment effect is not driven by people who correctly guessed the purpose of the experiment, people with high social desirability scores, or people for whom "transgender" was salient*

	Chose trans in private outcome round (=1)			
	Phases 1 + 2		Phase 1 only	
	(1)	(2)	(3)	(4)
3-person discussion	0.169***	0.164***	0.164***	0.133***
	(0.021)	(0.021)	(0.046)	(0.048)
	[<0.001]	[<0.001]	[<0.001]	[0.005]
Correctly guessed purpose (after main outcome)	0.193***			
	(0.039)			
	[<0.001]			
3-person discussion × Correctly guessed purpose (after main outcome)	−0.043			
	(0.061)			
	[0.478]			
Correctly guessed purpose (end of experiment)		0.064**		
		(0.032)		
		[0.048]		
3-person discussion × Correctly guessed purpose (end of experiment)		0.020		
		(0.052)		
		[0.695]		
Above median SDB score			−0.023	
			(0.033)	
			[0.486]	
3-person discussion × Above median SDB score			−0.014	
			(0.050)	
			[0.786]	
Transgender word remembered				0.043
				(0.039)
				[0.268]
Above median proportion of non-trans words remembered				0.000
				(0.025)
				[0.988]
3-person discussion × Transgender word remembered				0.031
				(0.054)
				[0.569]
Num. observations	4498	4498	2358	2358
Num. participants	2249	2249	1179	1179
Num. groups	751	751	393	393
Controls	X	X	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets, and use randomization inference for the *3-person discussion* coefficients. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm. Columns (1) and (2) include both phases 1 and 2. Columns (3) and (4) include only phase 1, when the SDB and salience modules were included. Only choices that include a transgender worker are included. The outcome is whether the participant chose the transgender worker in the private outcome round. *Columns (1) and (2).* Participants were asked what they thought the purpose of the study was twice: once after the main outcome round (column 1), and again at the end of the session (column 2). I class people as having correctly guessed the study's purpose if they say it is to measure preferences for hiring transgender individuals. *Column (3).* SDB score is the social desirability score based on the **?** index, described in Section **??**. *Column (4).* Participants were read two lists of words, described in Section **??**, and were asked to recall as many of the words as possible. *Transgender word remembered* indicates that the participant recalled the word "transgender". *Above median proportion of non-trans word remembered* indicates that the participant remembered more than 9 out of 17 of the other words in the two lists. Additional controls in all columns include: stratum fixed effects; phase fixed effects (for columns 1 and 2 only); dummies for rights videos; and controls selected by double LASSO (see Section **??**).

Table L14: *No significant differences in effect of the rights videos for participants who correctly guess the purpose*

	Chose trans in private outcome round (=1)	
	Phases 1 + 2	
	(1)	(2)
Rights messaging video	0.044**	0.049**
	(2.0×10^{-2}) [0.028]	(2.0×10^{-2}) [0.016]
Legal rights video	0.081***	0.079***
	(1.9×10^{-2}) [<0.001]	(2.0×10^{-2}) [<0.001]
Correctly guessed purpose (after main outcome)	0.049	
	(1.2×10^{-1}) [0.678]	
Rights messaging video × Correctly guessed purpose (after main outcome)	−0.021	
	(6.5×10^{-2}) [0.742]	
Legal rights video × Correctly guessed purpose (after main outcome)	−0.072	
	(6.3×10^{-2}) [0.255]	
Correctly guessed purpose (end of experiment)		0.031
		(4.5×10^{-2}) [0.500]
Rights messaging video × Correctly guessed purpose (end of experiment)		−0.047
		(5.4×10^{-2}) [0.386]
Legal rights video × Correctly guessed purpose (end of experiment)		−0.040
		(5.4×10^{-2}) [0.462]
Num. observations	6794	6794
Num. participants	3397	3397
Num. groups	1134	1134
Controls	X	X

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets, and use randomization inference for the *3-person discussion* coefficients. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm. Columns (1) and (2) include both phases 1 and 2. Columns (3) and (4) include only phase 1, when the SDB and salience modules were included. Only choices that include a transgender worker are included. The outcome is whether the participant chose the transgender worker in the private outcome round. *Columns (1) and (2)*. Participants were asked what they thought the purpose of the study was twice: once after the main outcome round (column 1), and again at the end of the session (column 2). I class people as having correctly guessed the study's purpose if they say it is to measure preferences for hiring transgender individuals.

Table L15: *Discussion effect is robust to increasing the stakes by offering 3 deliveries from the same worker*

	Chose worker in outcome round (=1)		Chose trans in outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.089**		
	(0.042) [0.038]		
3-person discussion	0.045	0.029	0.203***
	(0.032) [0.163]	(0.029) [0.318]	(0.048) [<0.001]
3 deliveries	0.030	0.035	−0.050
	(0.033) [0.374]	(0.028) [0.213]	(0.049) [0.300]
Worker is trans × 3-person discussion	0.167***	0.179***	
	(0.059) [0.005]	(0.059) [0.003]	
Worker is trans × 3 deliveries	−0.102*	−0.085	
	(0.060) [0.090]	(0.058) [0.147]	
3-person discussion × 3 deliveries	−0.019	−0.005	−0.054
	(0.044) [0.667]	(0.040) [0.905]	(0.074) [0.467]
Worker is trans × 3-person discussion × 3 deliveries	−0.022	−0.044	
	(0.086) [0.794]	(0.085) [0.608]	
Num. observations	3492	3492	1164
Num. participants	582	582	582
Num. groups	194	194	194
Controls		X	X
Controls interacted with worker is trans		X	

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the group-of-3 level and are in parentheses. randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes only the subsample of 582 individuals in phase 1 who were randomized into either receiving 1 delivery (N=288) or 3 deliveries (N=294). Participants who were offered 3 deliveries were (truthfully) told that they would receive 3 deliveries from the same worker, giving items of the same value each time. Phase 1 only included the *3-person discussion* arm and the *No discussion (private)* arm. Column (3) only includes choices that involved a transgender worker. In columns (1) and (2), the outcome is whether the *alternative worker* (rather than the male *benchmark worker*) is chosen in the private choices in the *outcome round*. In column (3), it is whether the transgender worker was selected. *Worker is trans* = 1 when the alternative worker is transgender, and is 0 when the alternative worker is male or female. The specification used is seen in equation **??**. Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; and the controls selected by double LASSO (see Section **??**). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown. Columns (2) and (3) also include controls for the relative # items offered by the alternative worker, the relative reliability score of the worker, and a dummy for whether the reliability score was shown.

Table L16: *Facilitator influence: more input from discussion facilitator about transgender choices is not associated with choosing transgender workers*

	Chose trans in outcome round (=1)	
	(1)	(2)
# words from facilitator, trans choices (Z)	−0.051**	
	(0.021) [0.014]	
# words from facilitator, non-trans choices (Z)	0.019	
	(0.021) [0.373]	
# words from resp., trans choices (Z)	0.017	
	(0.019) [0.376]	
# words from resp., non-trans choices (Z)	−0.063***	
	(0.021) [0.003]	
Prop. words from facilitator, trans choices		−0.160
		(0.133) [0.230]
Prop. words from facilitator, non-trans choices		0.256*
		(0.141) [0.070]
Num. observations	1638	1638
Num. participants	890	890
Num. groups	297	297

Notes: Sample only includes 3-person discussion participants. Word counts come from discussion transcript data. # *words* measures are converted to Z-scores so they have mean of 0 and standard deviation of 1. *Prop. words from facilitator* is the number of words from the facilitator as a proportion of words from both facilitator and respondents. Facilitators say 80 words on average about trans choices (43% of total words across facilitators and respondents); and 135 words about non-trans choices (51% of total). Outcome and controls are same as **??**, column 3.