Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India

Duncan Webb*

January 25, 2024

Job Market Paper

Download latest version

Abstract

Individual-level discrimination is often attributed to deep-seated prejudice that is difficult to change. But at the societal level, we sometimes observe rapid reductions in discriminatory preferences, suggesting that social interactions and the communication they entail might drive such shifts. I examine whether discrimination can be reduced by two types of communication about a minority: (i) *horizontal communication* between majority-group members, or (ii) *top-down communication* from agents of authority (e.g., the legal system). I run a field experiment in urban India (N=3,397) that measures discrimination against a marginalized community of transgender people. Participants are highly discriminatory: in a control condition, they sacrifice 1.9x their daily food expenditure to avoid hiring a transgender worker to deliver groceries to their home. But horizontal communication between cisgender participants sharply reduces discrimination: participants who were earlier involved in a group discussion with two of their neighbors no longer discriminate on average, even when making private post-discussion choices. This effect is 1.7x larger than the effect of top-down communication that informs participants about the legal rights of transgender people. The discussion's effects are not driven by virtue signaling or correcting a misperceived norm. Instead, participants appear to *persuade* each other to be more pro-trans, partly because pro-trans participants are the most vocal in discussions.

*Paris School of Economics, duncan.webb@psemail.eu. 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 Peter Andre, David Atkin, Luc Behaghel, David Bernard, Francis Bloch, Leonardo Bursztyn, Denis Cogneau, Oliver Vanden Eynde, Evan Friedman, Amory Gethin, Julien Grenet, Deivy Houeix, Nicolas Jacquemet, Eliana La Ferrara, Sylvie Lambert, Elisa Macchi, David Margolis, Edward Miguel, Oda Nedregård, Kate Orkin, Garima Sharma, Morten Nyborg Støstad, Nick Otis, Tom Raster, Kailash Rajah, Matthew Ridley, Gautam Rao, Chris Roth, Advik Shreekumar, Tavneet Suri, Simon Quinn, Jean-Marc Tallon, 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 field work. This work has been generously supported by JPAL's Crime and Violence Initiative, EUR-PgSE, CEPREMAP, the Development Research Group at PSE, the Weiss Fund, and the Institute for Humane Studies (grant # IHS017466). 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 both equity and efficiency in a wide range of economic domains, including in firms (Hjort, 2014; Glover et al., 2017; Hedegaard & Tyran, 2018), the labor market (Charles & Guryan, 2008; Folke & Rickne, 2022), housing (Christensen & Timmins, 2023), healthcare (Angerer et al., 2018), and informal social interactions (Lowe, 2020). Standard theories of discrimination focus on individuals' decisions based on beliefs or deep-seated preferences that are difficult to change (Becker, 1957; Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977). But at the societal level, we sometimes observe rapid reductions in discriminatory preferences, even without obvious changes in external conditions (Kuran, 1987; Fernández, 2013; Sunstein, 2019). 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] This suggests that some discrimination may be driven by malleable factors related to social interactions, such as social norms, a desire to conform, and communication between people.

In this paper, I therefore explore the idea that *communication* about a minority can generate rapid changes in discrimination, and can do so particularly through the interplay of norms and persuasion. First, I test whether generating communication between majority-group members about a minority ("*horizontal communication*") can affect discrimination. While theories of groupthink suggest that communication would amplify any existing discrimination (Myers & Lamm, 1976), communication could also reduce discrimination through a number of channels. For example, even when discrimination is common, it can be socially unacceptable to discriminate. People in group settings may therefore tilt their communication in favor of a minority in order to not be perceived as a discriminator. In doing so, they may convince each other to be less prejudiced—a change that can persist even after the group dissolves.[2] Another channel focuses on *who* communicates in a group: if those who are supportive of a minority are particularly vocal in a discussion, they may persuade others in a group to discriminate less. Second, I test whether communication about a minority from agents of authority ("*top-down communication*") can affect discrimination. One of the most powerful forms of such communication comes from the legal system: when minorities are granted legal protection, and this is communicated to citizens, this could act as a strong signal that discrimination is no longer socially acceptable, thus reducing discrimination (Sunstein, 1996; McAdams & Rasmusen, 2004; Benabou & Tirole, 2011). By contrast, when discrimination is institutionalized and perpetuated by the legal system (e.g., in apartheid regimes), the same

[1] Survey evidence shows that (i) the proportion of people in the UK indicating discomfort with interethnic marriage dropped from 55% in 1983 to 25% in 2013 (Park et al., 2014), (ii) the proportion of people in the US saying that homosexuality is wrong dropped from 80% in 1990 to 30% in 2020 (Center, 2014), and (iii) the proportion of people in Uganda saying that women should have equal rights went from 63% in 2002 to 80% in 2012 (Chingwete et al., 2014).

[2] The idea I explore here mirrors the logic of political correctness in other settings (Morris, 2001; Braghieri, 2021; Golman, 2022). If it is socially unacceptable to be openly racist or sexist, people may generate more favorable narratives about ethnic minorities and women, possibly leading to an equilibrium improvement in private attitudes towards these minorities.

mechanism could work to amplify discrimination among the populace.

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 people called *thirunangai*, primarily composed of transgender women. 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 (U.S. State Dept., 2021), and their distinct visual identity and historic role in Indian society make them highly recognizable – allowing me to measure discrimination through only the use of 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, discrimination also appears to be socially unacceptable,[3] raising the possibility that people may try to *appear* to be more pro-trans in a group setting and therefore persuade others to discriminate less. In addition, recent legal advances have affirmed that transgender individuals have fundamental rights, but awareness of these advances is low, allowing me to test whether communicating about legal rights can reduce discrimination.

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 reductions in subsequent discrimination: people who have earlier been involved in a group discussion discriminate substantially less. 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] 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 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

[3]For example, 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).

[4]Groups were always same-gender in order to avoid overly hierarchical relationships between group members. To make the discussions more naturalistic, and to make sure social image concerns played a role, we recruited neighbors who knew each other 98% of the time.

($p<0.001$), implying that anti-transgender discrimination is reduced to 0 on average (p=0.30). The effects are also partially persistent: when I re-survey participants approximately 1 month later, discussion participants are still 4 p.p. more likely to select transgender workers than the control group in a series of hypothetical hiring choices (p=0.03).

I then compare the effects of the discussion with top-down communication about the legal rights of transgender persons in India. Specifically, I cross-randomize whether participants are informed about a recent Indian Supreme Court ruling that affirmed that transgender people have all the same fundamental rights as other citizens, including freedom from discrimination.[5] Learning about these legal rights also lowers discrimination: participants are 10.3 p.p. more likely to select a transgender worker than a control condition ($p<0.001$). They also discriminate less than others who are shown persuasive messaging that advocates for the rights of transgender people without talking about the law (p of difference $\in [0.01, 0.12]$), suggesting that the legal authority of the Supreme Court may play some role. However, the effects of explaining the law are only 59% as large as the effects of involving participants in a group discussion (p of difference $\in [0.002, 0.04]$). The effects of information about legal rights do not persist when measured around 1 month later ($p \in [0.12, 0.51]$). In this context, therefore, allowing horizontal communication about transgender people is a more effective means of reducing discrimination than trying to actively reduce discrimination using top-down information.

I then seek to understand the mechanisms behind the effects, focusing on explaining the large impacts of the discussion (i.e., horizontal communication). Why does generating communication between privately discriminatory individuals sharply reduce post-communication discrimination? I show evidence against two candidate channels: (i) *correcting a misperceived norm*, and (ii) *virtue signaling*. Instead, the results appear to be driven by (iii) *persuasion*. People persuade each other to not discriminate, even in private after the discussion has ended, at least in part because pro-trans participants are the most vocal in the discussion.

First, I show evidence against the effects being driven by the process of *correcting a misperceived norm* (Bursztyn, González, & Yanagizawa-Drott, 2020). If participants initially overestimated how discriminatory their peers were, and the discussion corrected this misperception, then participants might have felt more comfortable choosing a transgender worker after the discussion. However, this pattern is not sufficient to explain the large treatment effects. In incentivized predictions of their fellow group members' private choices, participants in the control group do overestimate the extent of discrimination by 5 p.p. ($p<0.001$). But the predicted reduction in discrimination of 24 p.p. ($p<0.001$) generated by the discussion is far larger than the initial misperception. This suggests that a precisely corrected misperception could only account for 21% (95% CI: [8.9%, 32.5%]) of the discussion's treatment effect.

Second, I rule out that the discussion's effects are driven by a simple *virtue signaling* channel. If participants have social image concerns and do not want to *appear* discriminatory in a group

[5]There are no interaction effects between the legal rights video and the group discussion ($p \in [0.83, 0.96]$). 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.

setting, they may act more favorably towards transgender persons when making decisions that are visible to the rest of their group (Bénabou & Tirole, 2006; Bursztyn & Jensen, 2017). This could explain the discussion's effects if social image concerns encourage pro-trans behavior during the discussion that in turn persuades others to be more pro-trans. To test this channel, I use a *No discussion (public)* treatment arm in which participants do not discuss with each other, but instead make individual hiring choices that they know will later be revealed to other members of their group. If virtue signaling drives pro-trans behavior even in the absence of communication, this exogenous increase in social image concerns should reduce discrimination. Empirically, however, this treatment has no effect on average discrimination (p=0.46), suggesting that virtue signaling is not sufficient for explaining the effect of the discussion.

Third, I show evidence in favor of a *persuasion* channel. Participants appear to persuade each other with the narratives and justifications they share during the discussion, and overall they persuade each other to discriminate less because pro-trans participants are more vocal in the discussion. To test the presence of persuasion, I add a treatment arm in which one participant silently listens to two other people who have a 2-person discussion. The treatment effect on these "listeners" is just as large as on participants who actively participate in the 2-person discussion, both of which increase the probability of selecting a transgender worker by approximately 13 p.p. (p<0.001). This suggests that the effects are driven by persuasion *between* participants, rather than by participants persuading themselves or wanting to be consistent with their earlier actions.

To explain why participants are persuaded to be more pro-trans (rather than more anti-trans), I document suggestive evidence that the people who are most pro-trans are most vocal in the discussions. Each additional transgender worker chosen in the post-discussion choices (a proxy for pro-trans private attitudes)[6] is associated with a 32% higher probability of speaking first in the discussion (p=0.03) and a 27% higher probability of dominating the discussion (p=0.02), but *only* when discussing a choice that includes a transgender worker. In line with this, the overall pattern of communication during the discussions is highly pro-trans (for example, statements about transgender workers were 5.7x more likely to say something positive than to say something negative).

Motivated by this evidence, I present a model that describes the conditions under which pro-trans people would be more vocal in discussions, and how this can explain the large reductions in post-discussion discrimination. In the model, participants want to fit in with their group members: they can do so either by conforming to their group's preferences, or by changing their group members' preferences to match their own.[7] Because people who want to take a pro-trans action know that they will deviate further from the discriminatory preferences of their

[6] 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.

[7] The data provide evidence for this conformity motive: in the *No discussion (public)* condition, the intragroup correlation in participants' choices is higher (p=0.06). This suggests that participants try to match each others' preferences in a group setting.

group, they have a stronger incentive to persuade others to be more pro-trans, thus resulting in more pro-trans communication. The model shows that there can be a "sweet-spot" range, in which average preferences are discriminatory, but not *too* discriminatory, where *only* pro-trans participants try to persuade others, resulting in a reduction in post-discussion discrimination. Intuitively, average preferences must be discriminatory so that pro-trans people have a greater incentive to persuade others. They cannot be too discriminatory, otherwise no-one will take a pro-trans action at all, undermining the incentive to persuade others. Conversely, if average preferences are not discriminatory enough, anti-trans participants will speak up too, possibly making the discussion harmful overall. This provides a framework for thinking about the necessary conditions for horizontal communication to reduce discrimination.

Finally, I provide evidence against a number of alternative explanations of the results, including (i) increased attention or deliberation due to the discussion, (ii) social image concerns that affect even private, post-discussion choices, (iii) experimenter demand effects or social desirability bias, and (iv) salience.

This study makes four contributions. The key contribution is to show that discrimination can be rapidly reduced by generating organic horizontal communication about a minority, even in the absence of any additional information being injected into the group. This contrasts with standard theories in economics that attribute discrimination to deep-seated preference parameters that are hard to change (Becker, 1957), or to beliefs about minorities that only change when someone receives new information (Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977). While previous studies have shown that social contact between in-groups and out-groups can reduce discrimination (Allport, 1954; Pettigrew, 1998; Boisjoly et al., 2006; Pettigrew & Tropp, 2006; Paluck et al., 2019; Rao, 2019; Lowe, 2020; Corno et al., 2022), I show that even social contact *within* the in-group can reduce discrimination. Other work has also evaluated interventions where discussions are not peer-to-peer, but are driven by facilitators and designed to reduce discrimination (Bezrukova et al., 2016; Broockman & Kalla, 2016; Kalla & Broockman, 2020). By contrast, my study shows that the horizontal communication that endogenously arises within a group can reduce discrimination.

Second, I provide extensive evidence on the *mechanisms* that could explain why generating horizontal communication between privately discriminatory individuals lead to strong changes in discrimination, thereby contributing to a literature on social norms and social change (Kuran, 1987; Fernández, 2013; Sunstein, 2019; Gulesci et al., 2021; Andreoni et al., 2021). An important strand of literature has focused on cases of "pluralistic ignorance", in which there are misperceptions about the prevalence of discriminatory attitudes, suggesting that horizontal communication could correct these misperceptions and thereby change behavior (Kuran, 1987, 1991, 1997; Bursztyn, González, & Yanagizawa-Drott, 2020; Bursztyn, Egorov, & Fiorin, 2020).[8] In contrast, I show that even in the absence of such large misperceptions, social change can

[8]Other literature also examines ways of exogenously changing social norms in order to reduce discrimination (Dhar et al., 2022; Jayachandran, 2020; Gómez et al., 2018; Beaman et al., 2009; La Ferrara et al., 2012; Banerjee et al., 2019; Andrew et al., 2022).

be generated when people mutually persuade each other to change their attitudes, consistent with a model in which individuals self-select into being more vocal. I link to prior work on how people adapt their *communication* to conform to social norms, thereby promoting the spread of certain narratives (Braghieri, 2021; Morris, 2001; Golman, 2022; Crandall et al., 2002; Bénabou et al., 2020; Bursztyn et al., 2023). I show that such communication can be persuasive and therefore generate equilibrium changes in private behavior.

Third, I show that raising awareness of minority rights can reduce discrimination. This acts as an empirical validation of the *expressive law hypothesis*, that postulates that changes in the law may affect people's behavior by changing their perception of the prevailing social norm (McAdams, 2001, 2000; McAdams & Rasmusen, 2004; Benabou & Tirole, 2011; Sunstein, 1996). My work complements a recent empirical literature on how the law affects attitudes and norms (Lane et al., 2019; Funk, 2007; Aksoy et al., 2020; Chen & Yeh, 2014; Tankard & Paluck, 2017; Galbiati et al., 2020; Ofosu et al., 2019; Wheaton, 2020). I contribute by assessing whether communicating about the law can still affect behavior in a lower-middle income setting, where state capacity and trust in the legal system are lower. I also show that horizontal communication between individuals is more effective at reducing discrimination than top-down communication about the law.

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 (Badgett, 2014; Badgett et al., 2019), very little research in LMICs has examined its effects and causes (Badgett et al., 2021). There are notable exceptions in Latin America, including Gulesci et al. (2023) and Abbate et al. (2023), while Lyon (2023) shows that explaining to Ugandan citizens that homosexuality is legal in other countries leads to a backlash effect, worsening participants' opinions of those countries.[9] I contribute to this literature by examining a novel method of reducing discrimination through group discussions.

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*.[10] This group has a longstanding cultural and religious role in Indian society (Reddy, 2005; Kalra, 2012). Their visually recognizable identity, however, leaves them particu-

[9]Gulesci et al. (2023) shows that soap operas in Latin America with LGBTQ+ characters generate backlash, leading to more anti-LGBTQ+ attitudes. Abbate et al. (2023) examine discrimination in the housing market against LGBTQ+ individuals, showing that couples involving a transgender woman receive markedly fewer callbacks in their correspondence methodology, while gay male couples do not appear to face similar discrimination. Research on anti-LGBTQ+ discrimination in Europe and the US has sought to understand its magnitude and nature (Tilcsik, 2011; Carpenter et al., 2020; Flores, 2015; Drydakis, 2022; Klawitter, 2015; Burn, 2020; Button et al., 2020; Granberg et al., 2020),along with whether discrimination can be reduced by changes in the law (Sansone, 2018; Aksoy et al., 2020; Tankard & Paluck, 2017; Ofosu et al., 2019) or information interventions (Aksoy et al., 2021). Most relevant to the current study, Granberg et al. (2020) show some evidence of hiring discrimination against transgender individuals in an audit study in Sweden.

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

larly susceptible to discrimination (Sharma, 2014; Agoramoorthy & Hsu, 2015).

The Indian Census (2011) estimates there to be at least 490,000 transgender people in India, but given their marginalized status, the actual number is likely to be much higher (Dixit et al., 2023). Anti-transgender discrimination is therefore likely to result in substantial welfare and efficiency costs (Badgett, 2014).

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 (Masih et al., 2012; Shivakumar & Yadiyurshetty, 2014; Badgett, 2014; Nuttbrock, 2018). In line with this, a survey in north India indicated that only 6% are involved in formal employment (Kerala Development Society, 2017). Discrimination can also take other forms, such as being cut off from family support, housing discrimination, harassment, violence, and difficulty in accessing medical treatment and education (IPSOS, 2018; U.S. State Dept., 2021; Mal, 2015; Ganju & Saggurti, 2017; Shaikh et al., 2016; Baba & Sogani, 2018; Chakrapani et al., 2004; V. Chakrapani et al., 2011; Sangama, 2015).

Widespread prejudice and discrimination may be increasingly at odds with social norms that penalize discriminators. Discrimination, though common, can generate social disapproval: for example, in the study control group, discriminatory scenarios were rated as "wrong" 93% of the time by respondents.[11] This aligns with other survey evidence that indicates widespread support for protecting transgender people from discrimination (IPSOS, 2018). 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 (Bursztyn, Egorov, & Fiorin, 2020; Bursztyn et al., 2023).

Legal changes may have contributed to the decreasing social acceptability of discrimination. In 2014, 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 (see more detail in Appendix E). By institutionalizing transgender rights, these changes may reduce discrimination in social settings by signaling to the populace that discrimination is no longer socially acceptable. At the same time, awareness of these recent changes remains low. In the study's control condition, 36% either believes that trans individuals do not have any legal status in India, or cannot identify a single legal right they hold, allowing me to change participants' beliefs about the law and examine the effects on discrimination.

The study took place in Chennai, the largest city in the state of Tamil Nadu. Tamil Nadu is an appropriate context for studying anti-transgender discrimination, because despite seeing policy changes that favor transgender people (e.g., the state government constituting a Transgender Welfare Board), qualitative studies indicate that discrimination is persistent and widespread

[11] These self-reported ratings may be vulnerable to social desirability bias. But such bias will lead people to report what they believe to be the socially appropriate answer, in line with the claim that social norms curtail the expression of prejudice.

(V. Chakrapani et al., 2011; Delhi, 2018; Kumar et al., 2022; Subramanian et al., 2009). The urban setting is also advantageous for two reasons. First, our scoping work suggested that urban residents were more likely than rural residents to have recently seen or interacted with a transgender person, so awareness and visual recognition of transgender people is high. Second, urban residents are familiar with online delivery services, allowing us to use delivery service market research as a plausible framing 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) and top-down communication (information about transgender rights) on the subsequent level of hiring discrimination against transgender workers.

The main goals of the experimental design were: (i) measuring discrimination in a realistic setting, with choices that had real stakes for participants; (ii) generating natural horizontal communication between participants about transgender people, without it being obvious about the purpose of the study; (iii) delivering information about the legal rights of transgender people while minimizing experimenter demand effects; and (iv) understanding the mechanisms underlying treatment effects, such as the role of social norms and persuasion.

All treatments and the primary data collection took place over the course of a single session that lasted approximately 1 hour. To allow for a group discussion, enumerators always recruited and then interviewed 3 respondents at the same time. I call these 3 respondents a "group".

To measure hiring discrimination against transgender workers, I offered participants a free grocery delivery to their home, and asked them to make a series of choices over the worker who would carry out the delivery, and which items they would receive. Each participant made 10 binary choices, one of which was randomly selected to be implemented. Between 2 and 9 weeks after the main session, the selected delivery option was carried out by the chosen worker, and participants were asked follow-up survey questions.

To measure the effects of a group discussion, the first 4 choices (the *treatment round*) were used as a source of treatment variation. The remaining 6 choices (the *outcome round*) were always made individually and in private. It is these later private choices that I use as my main outcome, allowing me to examine the effect of the treatments on private individual choices.

The first set of treatments were designed to measure the effects of horizontal communication between participants, and the mechanisms underlying such effects. Each treatment corresponds to a different process for the treatment round. The main effect measured by the comparison

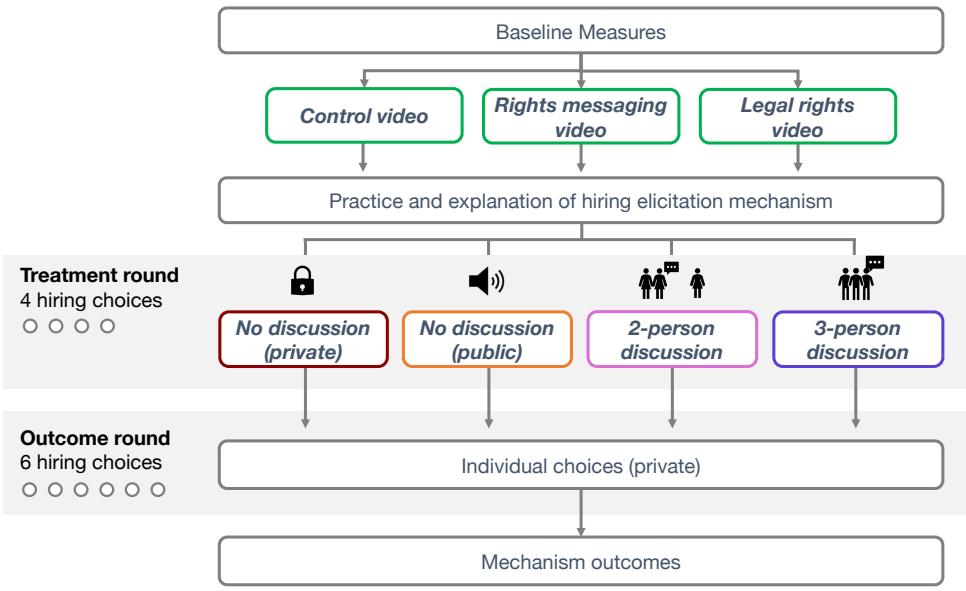
[12] 80% of the sample say that they have previously ordered goods to be delivered to their home using an app, reflecting the popularity of meal delivery services such as Swiggy and Zomato. The market research framing is also not so unusual in this context: 29% of the sample have previously taken part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

between (i) *3-person discussion*, in which all 3 participants had a discussion about their preferred hiring options and made collective choices; and (ii) *No discussion (private)*, a control condition in which all participants made private individual hiring choices. Then, two further treatment arms were used to understand the mechanisms behind the effects of horizontal communication: (iii) *2-person discussion*, in which 2 participants in the group had a discussion and made a collective choice, while the third participant listened; and (iv) *No discussion (public)* in which all participants made individual hiring choices that they knew would later be revealed to the others in their group.

The second source of treatment variation was designed to test the effects of top-down communication about legal rights. Specifically, I cross-randomized a video shown to participants before they make 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 transgender rights, or (iii) a *control* video that did not mention transgender rights.

Figure 1 shows a summary of the experimental design for the main session (Figure A1 gives further detail). I describe the design in more detail below.

Figure 1: *Summary of experimental design*



Notes: More detail is given in Appendix Figure A1.

3.2 *Sample and recruitment*

Participants were recruited from urban areas in Chennai between March and July 2023 (see Figure A2 for survey locations). They were recruited through direct household canvassing and introductions from community leaders. All participants were aged 20-65 and could read Tamil.

To allow for the group discussion, enumerators always recruited and then interviewed 3 respondents at the same time. To avoid recruitment strategies that differed across treatments,

enumerators were blind to treatment status before starting the survey. This means that even participants in the control group were recruited as a group of 3. All 3 members of a group were interviewed simultaneously.

To make any group activities as naturalistic as possible, all members of a group were neighbors or acquaintances that lived on the same street or within the same locality. The 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, and we did not recruit multiple members of the same household in a group. The majority of the sample (85%) was female. The framing of the study as a market research survey about deliveries was more relevant for females, since they were more likely to be generally responsible for managing household food expenditures and receiving deliveries (88%) than men (59%).

3.3 Hiring choices

To obfuscate the purpose of the study and reduce experimenter demand effects, the survey was framed as a market research survey, and participants were truthfully told that we were trying to understand people's preferences for grocery delivery options. After the main hiring choices, only 8% of the sample had correctly guessed that the purpose of the experiment was related to transgender workers (and the treatment effects are not driven by these participants, see Section 7).

All participants made a series of 10 binary choices over which delivery option they preferred, one of which was randomly selected to be implemented. Figure 2 shows an example of one such binary choice. For each choice, participants saw two options. Each option always included a photo of the worker and the items on offer, in some cases inducing a trade-off between a preferred worker and preferred items.[13]

Some workers were transgender, and participants were able to visually recognize them as such. In a supplementary study (N=114), carried out between August and September 2022, participants correctly identified transgender worker photos as being transgender 97% of the time (Appendix Table A3).

To ensure the participants anticipated some social contact with the worker, they were truthfully told that they would have a 15-minute conversation at their home with the selected worker when the delivery took place, during which they would be asked questions about their satisfaction with the service. When introducing the hiring process, participants were instructed to consider the worker and their characteristics, the items they offered, and this 15-minute conversation.

In each choice, there was a "*benchmark*" option who was cisgender male, and an "*alternative*"

[13]To minimize noise generated by differences in photos, all photos were headshots with a neutral grey background in which the worker had a neutral expression.

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

option who was either cisgender male, cisgender female, or transgender.[14] Throughout the paper, I measure anti-transgender discrimination as the reduction in the probability that the *alternative* worker was chosen when the alternative was transgender. For simplicity, I call this person the "worker". The alternative worker was transgender for 4 choices, implying that 20% of the 20 photos they saw were transgender. This proportion was chosen to ensure sufficient power without making the purpose of the experiment too obvious to participants. The position of the alternative (left or right), the order of choice-pairs, and the selection of worker photo for a given gender were all randomized. Across the main hiring choices, the same worker was never seen twice by a respondent.[15]

To evaluate how participants traded off material benefits with their preferences for workers, the number of items offered by each worker was randomly varied so that sometimes one worker offered more items than the other. This randomization was balanced across worker genders. Each option either offered 1 item (masala spice mix), 2 items (masala and tea), or 3 items (masala, tea, and ghee).[16] 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: they cost Rs. 68, Rs. 154, and Rs. 240 respectively, corresponding to 103%, 234%, and 365% of median daily per capita food expenditure.

[14]This set-up reduces the number of gender combinations, thereby increasing power on the male-to-trans comparison, although it does not allow me to directly measure preferences between trans and female workers.

[15]For the main hiring elicitation, photos were selected from a pool of 20 cisgender males, 21 cisgender females, and 13 transgender people. The cisgender photos were of survey enumerators recruited to do the main survey. Participants who were interviewed by a specific enumerator team were not shown pictures of that same team to avoid response bias. The transgender photos were of workers who consented to the use of their photo and who agreed to carry out deliveries if they were selected. For later mechanism outcomes (that did not require a worker to actually deliver goods), I used a set of stock headshot photos.

[16]The randomization 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.

In some choice-pairs, additional truthful signals of worker quality were shown. These were included in order to evaluate the extent to which discrimination against transgender workers was *statistical*, i.e., driven by beliefs about whether they would reliably complete a delivery. Some choice-pairs reported the true proportion of successful deliveries from a timed training exercise carried out by all workers (the "reliability score"). Participants were told that this was the proportion of completed deliveries from a training exercise. Workers completed more than one training exercise with different time limits, and I randomly showed their score based on one of three categories: their low score (5 or 6), their mid-value score (7 or 8), or their high score (9 or 10). This yielded exogenous variation in the perceived quality of each worker (see Appendix D for discussion of the ethical considerations). 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.

Implementation of choices. To ensure incentive-compatibility, one of the participants' 10 choices was randomly selected to be implemented using scratch-cards, and the participant received a delivery from the chosen worker 2–9 weeks after the main session. At the time of the delivery, the worker carried out a short follow-up survey. To minimize risk to transgender workers, the randomization was designed so that choice pairs that included a transgender worker were selected by the scratch-cards in fewer than 1% of cases.[17]

To ensure the participant understood the randomization scheme, participants first took part in a practice round, in which they made a series of 4 binary choices between items worth less than Rs. 5. Mimicking the main hiring elicitation, the enumerators used a scratch-card to select which of the 4 choices was actually implemented. We also asked a series of comprehension checks before the practice round and main hiring round, and re-explained to respondents if they answered incorrectly. Participants responded correctly to these questions the first time they were asked 92% of the time in the practice round, and 86% in the main hiring round, suggesting a high level of comprehension overall.

3.4 Treatments

3.4.1 Discussion arms

To measure the effects of horizontal communication (in the form of a group discussion), I varied the elicitation process for the first 4 hiring choices (the *treatment round*). In this round, 2 of the 4 pairs included a transgender worker. All three participants in the same group always saw the same delivery options, regardless of treatment status.

In the treatment round, groups were randomized into one of the four conditions described

[17] Transgender workers doing deliveries could have been vulnerable to stigma and abuse. The randomization was designed to avoid this as much as possible, while also truthfully telling participants that they could receive a delivery from any of the workers they chose. For the few transgender workers who actually carried out a delivery, the worker was accompanied by a team of 2-3 enumerators and a supervisor, and interaction between the transgender worker and participant was reduced to a minimum (see Appendix D for discussion of the ethical considerations).

below. randomization was stratified by participant gender and survey team.

1. **3-person discussion** (N=890). Respondents took part in a discussion among their group of 3 neighbors, in which they discussed which workers they preferred and why, and then make *joint* choices. I describe the design of this discussion in more detail below.

2. **2-person discussion** (N=549). 2 participants (the "*speakers*") were randomly selected to take part in a discussion. The 3rd, the "*listener*", was asked to stay silent and simply listen to the speakers' choices and justifications. The listener was a mechanism treatment designed to gauge the impact of listening to a discussion without taking part in one oneself.

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 arm was a mechanism treatment designed to evaluate how participants' choices were affected by social image concerns in the absence of a discussion. In addition, I varied the timing of the announcement. This allowed me to see if simply being told that another person had selected a transgender worker could reduce subsequent discrimination, and how this compared to hearing a discussion about a transgender worker. 2 randomly-selected participants out of 3 (the "*observers*") were told others' choices *before* making their private outcome-round choices, allowing me to measure the persuasive effect of observing others' choices. The 3rd participant (the "*non-observer*") was only told *after* making their private outcome-round choices. 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 their treatment round choices.

4. **No discussion (private)** (N=1365). The treatment round was answered individually and in private. This was a control condition to act as a comparison to other treatments.

Enumerator observations suggest that participants correctly followed the protocols.[18]

3.4.2 Design of the discussion

For the *3-person discussion*, *2-person discussion* and *No discussion (public)* arms, participants completed the treatment round together in their group of 3. The group activity usually took around 10 minutes. The activity usually took place one of the participant's homes, or in another nearby common area (e.g., the common courtyard in a tower block). For the *No discussion (private)* treatment, participants were interviewed separately, out of earshot from one another.

In the *2-person* and *3-person discussion*, discussion participants had to reach a *collective* decision

[18] *Listeners* did not say anything about the options for any choice in 93% of groups, according to enumerator observations. To make social image concerns salient, *No discussion (public)* participants chose in group setting. To ensure participants did not influence each other *during* the elicitation process, they were told to not 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.

for each pair. If the scratch-cards selected one of these pairs, each member of the group received the same bundle of items from the same worker. To ensure naturalism, participants were truthfully told that it would be logistically easier for us to organize the same worker to deliver the same items. In contrast, those in the *No discussion* arms simply selected the option they preferred individually.

In each discussion, respondents discussed their opinion of each option, explained why they preferred one option or another, and tried to convince the group to choose their preferred option in cases of disagreement. To minimize demand effects, and ensure that the discussion involved *horizontal communication* that naturally arose between participants, the enumerator leading the discussion never mentioned the word *transgender* themselves. Instead, any discussion of transgender people was only initiated by the participant's response to a photo they saw (see Appendix H for the discussion script and further details).

3.4.3 Rights videos

To test the effect of top-down communication about minority rights, I cross-randomized a second set of treatments. Participants were shown one of three different videos about rights before making their hiring choices. All videos lasted between 80 and 90 seconds, and were narrated in Tamil by a local member of the research team (who was not shown). The majority of the content was the same across all three videos, and explained consumer and worker rights in the context of delivery services, in line with the framing of the study as a market research survey for a delivery service. As treatment variation, I varied one of the examples used when explaining what "rights" were:

1. **Legal rights video** (N=1135). Participants were told that transgender people have *legally instituted* rights in India. This video was designed to measure the effect of changing people's beliefs about the law on the level of discrimination. Specifically, they were told: *"As another example, 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). Participants were told that transgender people *should* have rights, but they were not told that they legally *do* have those rights. This was intended to designed to measure whether legal protection is important for reducing discrimination, or if simply communicating a narrative about the transgender rights without institutional support is sufficient. The wording was kept as similar as possible to the legal rights video: *"As another example, 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 were not given information about the rights of transgender people. Instead, the video included placebo information about voting rights: *"As another example, 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."

Appendix I contains the full video scripts. To ensure participants could hear 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.

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 was divided into two phases (see Figure 3). Phase 1, completed between March and April 2023, included only the *No discussion (private)* and *3-person discussion* arms. In phase 2 (May–July 2023), the *No discussion (public)* and *2-person discussion* arms were added. These arms were added upon the receipt of additional funding, and were designed to understand the mechanisms behind the effects of the 3-person discussion arm. In order to be able to detect small effects on the mechanism treatments relative to the control group (*No discussion, private*), I added additional sample size to the control arm in phase 2.[19]

When analyzing the data, I primarily make use of three different samples:

1. **3-person discussion sample** includes both phase 1 and 2 of the *No discussion (private)* (N=1365) and the *3-person discussion* (N=890) arms. It is used to measure the effect of the 3-person discussion.

2. **Phase 2 sample** uses only phase 2, and includes all treatment arms. This is used to analyze the effect of the mechanism treatments relative to the *No discussion (private)* arm and the *3-person discussion* arm.

3. **Video sample**. Since the rights videos are cross-randomized across all discussion arms in both phases, I use all data from all phases and all discussion arms when analyzing the effects of the videos. I also control for the phase of data collection for this sample.

[19]This creates an imbalance: control-group observations are relatively more likely to come from phase 2 than *3-person discussion* observations. I control for phase fixed effects in all relevant specifications that include controls. The results are also robust to adding sampling weights that re-balance the treatment conditions (Table A4).

3.6 *Pre-analysis plan*

I preregistered the design of both phases, and document any deviations from the pre-analysis plan in Appendix G. I changed the main specification to exclude interaction terms between discussion treatments and video treatments in order to ease interpretation and increase power (although I also show interacted specifications in Table A12 and Figure A13). The other main deviations were primarily due to unexpectedly low survey productivity in phase 1 that led to tighter-than-expected budget constraints. These include: (i) dropping a plan to include a mixed-video arm in which participants in a group saw different rights videos; (ii) carrying out deliveries after 2–9 weeks instead of 1 week; and (iii) reducing the number of mechanism outcomes.

3.7 *Balance checks*

For all samples and relevant treatment comparisons, the treatment groups were well-balanced across key characteristics (Tables A5, A6, and A7). For each arm, a joint F-test that compares it to the control condition indicated no systematic differences in observable characteristics. 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 (Table A5), and *rights messaging video* participants came from slightly larger households, with a slightly lower per capita food expenditure (Table A7). I use LASSO to select all controls that predict both treatment status and outcomes (as per Belloni et al. (2014), see Appendix J.9), so these imbalances are unlikely to affect my results.

3.8 *Outcome and specification*

The pre-specified primary outcome is participants' choices in the outcome round of hiring, which took place *individually* for all treatment groups. The design thereby aims to estimate the causal effect of discussion and the rights video on participants' *subsequent* private choices.

The outcome round choices were designed to be private, so participants who had previously been in a group setting moved to be out of earshot of one another. Accordingly, 94% of respondents reported that others in their group could *not* hear their responses in the outcome round.

However, the choices were arguably not *completely* private, because (i) participants knew each other and might ask each other what they chose, (ii) they might anticipate that neighbors would observe the delivery worker when the delivery took place, and (iii) enumerators observed the answers given by respondents. These imply that social image concerns may still play a role in the outcome round choices. While I cannot rule out channel (i), for robustness I design an "extra private" outcome that addresses channels (ii) and (iii) (see Section 7).

The outcome round included 6 binary choices in the same format as the treatment round, two of which included a transgender worker. The main specification for participant i in group j,

when making a choice for the pair of workers k, is:

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

where:

- $ChooseAlternative_{ijk} = 1$ 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 in pair k shown to i is a transgender individual, 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 in group j is in treatment arm τ, with $\mathcal{T}$ being the set of treatments analyzed. These are either (i) a dummy for the 3-person discussion, (ii) dummies for each discussion-arm treatment, or (iii) dummies for each rights video. I do not include interaction effects between the videos and discussion arms in the main specification, but they are shown in Table A12 and Figure A13.
- $\mathbf{X}_{ijk}$ 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 the reliability score, the benchmark worker's reliability score, an indicator for 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 (Belloni et al., 2014) to select an additional set of controls that predict both the treatment and outcome variables (see Appendix J.9 for the variables used).

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

The main treatment effect is measured by the coefficients β_τ, which describes the reduction in discrimination caused by the treatment. When interacted controls are not included, γ describes the baseline level of discrimination against transgender workers in the hiring choices in the relevant control group. Standard errors on regressions are clustered at the group-of-3 level. For tables in the main text, I use randomization inference to calculate p-values (Young, 2019). Since I have only one primary outcome, I do not correct it for multiple hypothesis testing.

3.9 Mechanism outcomes

In addition to the main outcome variable, I elicited other measures designed to understand the mechanisms behind the results (see Appendix J and the relevant results section for more detail on each). Some measures were included only for a single phase of data collection, as I specify below.

Baseline measures. We elicited several baseline measures, including (i) demographics; (ii) susceptibility to social desirability bias based on Crowne & Marlowe (1960) (phase 1 only); (iii) questions about the proximity of relationships between group members (phase 2 only); and (iv) a persuasiveness index designed to measure how persuasive an individual was likely to be in a discussion (phase 2 only). These questions were intermingled with questions about deliveries, in order to reinforce the framing of the study as a market research survey.[20]

Treatment round choices. I use the hiring choices during the *treatment round* as a pre-specified secondary outcome. This allows me to examine what choices were made *during* (rather than *after*) the discussion, and compare this to the control group's individual, private choices.

Group observations. During the group activities, one enumerator facilitated the discussion, instructing the participants on what to do and prompting participants. Another enumerator marked a series of observation questions about the group activity, which were pre-specified as secondary outcomes. For example, in the discussion arms, they marked who spoke first, who dominated the discussion, the main reasons participants cited in the discussion for making their choices, who spoke in favor of a transgender option, whether anyone said something positive or negative about transgender workers, and how much discussion occurred for each pair.[21]

Post-hiring mechanisms. Immediately after the hiring choices were completed, we elicited a series of further mechanism outcomes. As pre-specified secondary outcomes, I included: (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 (Droitcour et al., 2004; Glynn, 2013) measuring the proportion of people agreeing with a discriminatory statement; and (v) questions about the legal status of transgender people (along with other questions about the rights of delivery workers to obfuscate the purpose of the section). Then, as exploratory mechanism checks, I include: (vi) hiring choices for a private grocery pick-up involving interaction with a worker (phase 2 only); (vii) recall checks, in which participants were asked to recall the choices made by

[20] I did not include baseline measures of attitudes towards transgender people, 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, evidence from behavioral economics suggests that people may have a desire to be or appear consistent with previous actions (Falk & Zimmermann, 2017), or may persuade themselves to make their preferences align with their previous actions (Schwardmann et al., 2022). If true, eliciting baseline attitudes or discrimination would anchor people's behavior to a pre-treatment state, and lead treatment effects being underestimated. Second, 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 towards transgender people, contrasted to the hiring questions that are subtler and less obviously focused on discrimination.

[21] We also asked for consent to record the audio of the discussion. Consent was refused in 16% of discussions. Table A8 shows that the treatment effects of the 3-person discussion are not significantly different for groups that refused and consented to the audio recording.

themselves or others earlier in the survey (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 was carried out, an average of 35.3 days after the initial survey (SD: 14.4 days), we elicited a short (15 minute) survey to measure how persistent treatment effects were. As a pre-specified secondary outcome, we asked 6 more hypothetical hiring choices with a new set of worker photos, and a different set of grocery items. We made clear to respondents that these 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).

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 1 and Figure 4). 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 1, 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 include a transgender worker (Column 3), to using a logit rather than a linear probability model (Table A9), and to dropping the 6% of cases where the outcome round was overheard by neighbors (Table A10). The treatment effect of the discussion is also robust and still significant at the 0.1% level when restricting the sample to participants who did not see a video discussing transgender rights (Table A11). The main effects thus hold even for participants who do not receive any information about transgender rights, suggesting the effects are not driven by interactions between the rights videos and the group discussions.[22]

To benchmark the size of the reduction in discrimination, I use the random variation in the items offered across the options in a pair to infer the average willingness to pay to avoid

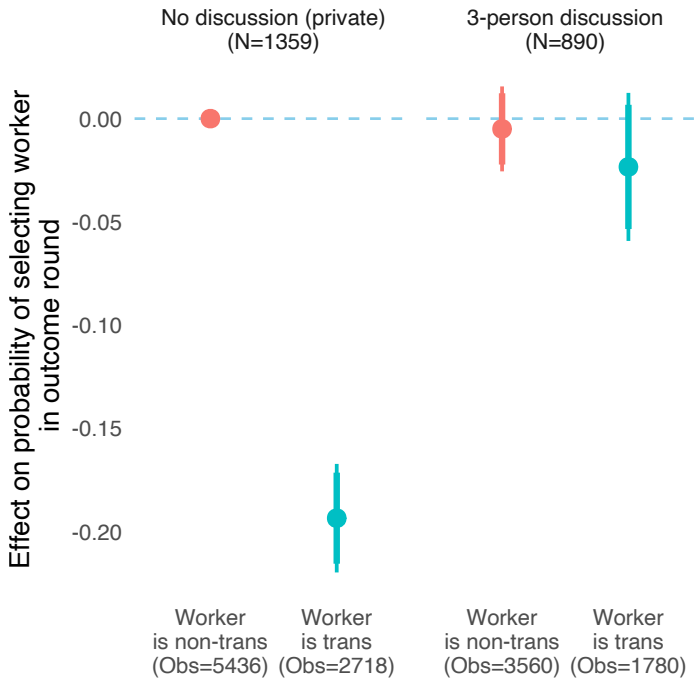
[22] Recent work in econometrics suggests that when using cross-randomized designs, regressions that do not account for interaction terms can yield incorrect inference (Muralidharan et al., 2023). There is no evidence of interactions between the *legal rights* video and discussions (Table A12), but weak evidence of a negative interaction effect between the *rights messaging* video and the discussions, suggesting that these two interventions are substitutes. Throughout the paper, I present results that control for the all treatments. The coefficients on the discussion should therefore be interpreted as conditional on the distribution of the other video treatment. However, given that 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]	(0.022) [<0.001]	
Worker is trans	−0.193***		
	(0.013) [<0.001]		
3-person discussion	−0.004	0.001	0.167***
	(0.011) [0.716]	(0.010) [0.912]	(0.020) [<0.001]
Relative # items offered		0.128***	0.097***
		(0.005) [<0.001]	(0.008) [<0.001]
Relative reliability score		0.017***	0.013**
		(0.003) [<0.001]	(0.005) [0.012]
Reliability score is shown (=1)		0.022***	0.040***
		(0.008) [0.007]	(0.012) [0.001]
Num. observations	13 494	13 494	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Mean: no discussion (private), worker is non-trans	0.61	0.61	
Mean: no discussion (private), worker is 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*) 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 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. The specification used is seen in equation 1. 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 J.9). 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.

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



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

choosing a transgender worker (Figure A14). In the *No discussion (private)* arm, participants are on average willing to sacrifice items worth Rs. 127 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 willingness to pay to avoid is Rs. 13 (p of difference $<$ 0.001), and is no longer significantly different from 0 (p=0.265). In Appendix B, I structurally estimate a model that allows for preferences to be correlated within participants and groups, and comes to similar conclusions. The results therefore suggest that the discussion generates a large 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 (Table A15, Figure A16). In the *No discussion (private)* arm, even when shown a transgender worker that is weakly dominating on items and reliability score, participants still select the non-transgender worker 47% of the time. By contrast, in the *3-person discussion* arm, this figure has reduced to 29% (difference: 17.9 p.p., p<0.001).

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 (Table 1, column 2). And people

were less sensitive to items when shown a transgender person (Table A17, column 1), a result that holds for both treatment conditions (Table A17, column 2).[23]

Belief-based (statistical) discrimination appears to underly some of participants' unwillingness to select transgender people, driven by negative stereotypes that portray transgender workers as unreliable. Despite transgender workers having the same average reliability score as other genders, participants rate transgender workers as less likely to complete a delivery (Table 4, 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 (Table A18, column 1). Anti-transgender discrimination in the control group therefore appears to be partially driven by a form of inaccurate statistical discrimination (Bohren et al., 2023).

However, the effect of the discussion does not appear to be based on changes in such statistical discrimination. The discussion does not significantly affect beliefs about the reliability of transgender workers (Table 4, 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 (Table A18, column 2). 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).

A heterogeneity analysis (Table A19) shows that while anti-transgender discrimination is stronger for male participants than female participants (difference: 6.8 p.p., p=0.07), the *treatment effects* of the discussion are similar for both males and females (p=0.95). 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 women (estimate: 3.4 p.p., p=0.12).[24]

Finally, the discussion consistently reduces discrimination across all 13 transgender worker photos used (Figure A21). This suggests that other features of the worker photos do not drive the results. For example, suppose the transgender workers tended to appear to be poorer, and the discussion actually increased preferences for choosing poor people. We would nevertheless expect at least *some* transgender workers to appear richer, leading to a coefficient estimate in

[23]The fact that the sensitivity to items did not vary across treatment conditions (Table A17, columns 3 and 4) 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.

[24]There is heterogeneity in levels separating the analysis of non-transgender workers into males and females (Figure A20). 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).

the opposite direction for a subset of photos – something that is not observed empirically.

4.2 *Effect of transgender rights videos*

The results on the videos about transgender rights show that they also reduce discrimination in the outcome round, although significantly less than the discussion (Table 2). Both the *Rights messaging* video and the *Legal rights* video led to significant increases in the probability of selecting a transgender worker in the outcome round, with coefficients of 5.8 p.p. and 10.3 p.p. respectively. There is some evidence that the legal rights video has a stronger effect than the rights messaging video ($p \in [0.01, 0.12]$, depending on the specification). Endorsing transgender rights thus appears to reduce discrimination more effectively when it is backed by the legal authority of the Supreme Court.[25] 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 $\in [0.002, 0.04]$).

I do not find interaction effects between the *legal rights* video and group discussions (Table A12 and Figure A13); the reductions in discrimination caused by both combine approximately linearly ($p \in [0.83, 0.96]$). By contrast, there is weak evidence of a negative interaction effect between the *rights messaging* video and group discussions, so that the rights messaging video has no detectable effect on discrimination in the group-discussion arms ($p \in [0.80, 1.00]$). 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.

4.3 *Persistence of effects*

To examine whether the effects of the discussion and the rights videos persist over the medium-run, we elicited a follow-up survey when the delivery was carried out. This survey took place an average of 35.3 days after the initial survey (SD: 14.4 days). 95.7% of the sample were found,[26] and there was no evidence of differential attrition (Table A23). 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 elicited 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 on these hypothetical choices that were still present after 2-9 weeks (Table 3, panel A). Participants were approximately 4 p.p. more likely to select transgender workers in the hypothetical follow-up choices (p=0.03). By contrast, the videos about transgender rights did not lead to a detectable persistent effect

[25] 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 Table A22).

[26] 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.

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) [0.224]	(0.010) [0.105]	(0.019) [0.008]
Legal rights video	−0.019*	−0.022**	0.081***
	(0.011) [0.078]	(0.010) [0.029]	(0.019) [<0.001]
Worker is trans × Rights messaging video	0.058***	0.070***	
	(0.023) [0.001]	(0.022) [<0.001]	
Worker is trans × Legal rights video	0.103***	0.104***	
	(0.022) [<0.001]	(0.020) [<0.001]	
Num. observations	20 382	20 382	6794
Num. participants	3397	3397	3397
Num. groups	1134	1134	1134
Mean: control video, worker is non-trans	0.62	0.62	
Mean: control video, worker is trans	0.45	0.45	0.45
Controls		X	X
Controls interacted with worker is trans		X	
p(Rights messaging video=Legal rights video)	0.012	0.045	0.122
p(Legal rights video=3-person discussion)	0.024	0.040	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 Table 1.

on discrimination ($p \in [0.12, 0.51]$, Table 3, panel B). Since these choices are hypothetical, the results are more vulnerable to concerns about experimenter demand effects and social desirability bias. In line with this concern, the control group discriminated less in the hypothetical follow-up survey than in the main survey (12.7 p.p., difference: 7.1 p.p., p<0.001).[27] However, participants' hypothetical choices in the follow-up do positively correlate with their incentivized choices in the main survey (ρ=0.32, p<0.001). The results therefore suggest that the large short-term effects of the discussion on discrimination may translate into medium-run effects, while the information about transgender rights was not sufficiently impactful to have a medium-run effect. This raises the possibility that even short interventions involving horizontal within-group communication could have persistent effects on behavior.

[27]There is also a risk that participants from different treatment groups communicated with each other after the main survey. However, I find no evidence of geographical spillovers (Table A24).

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

Panel A: Effect of 3-person discussion (3-person discussion sample, phases 1 + 2)

	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) [0.776]	(0.011) [0.631]	(0.019) [0.026]
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
Mean: no discussion (private), worker is non-trans	0.62	0.62	
Mean: no discussion (private), worker is trans	0.49	0.49	0.49
Controls		X	X
Controls interacted with worker is trans		X	

Panel B: Effect of transgender rights videos (all participants)

	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) [0.781]	(0.011) [0.792]	(0.020) [0.447]
Legal rights video	−0.002	−0.001	0.028
	(0.011) [0.863]	(0.010) [0.951]	(0.019) [0.159]
Worker is trans × Rights messaging video	0.012	0.012	
	(0.021) [0.506]	(0.021) [0.524]	
Worker is trans × Legal rights video	0.029	0.028	
	(0.021) [0.117]	(0.021) [0.106]	
Num. observations	19 266	19 266	6416
Num. participants	3230	3230	3224
Num. groups	1134	1134	1133
Mean: control video, worker is non-trans	0.62	0.62	
Mean: control video, worker is 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 in panel A includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phases 1 and 2. Sample in panel B includes all participants. Controls in panel A include dummies for the rights videos, and controls in panel B include dummies for the discussion-arm treatments, as well as the other controls specified in Tables 1 and 2. In the follow-up survey, workers in a pair always had the same reliability score and offered same number of items. Specification is otherwise the same as Tables 1 and 2.

5 Intermediate outcomes: effects on attitudes, beliefs, norms, and behavior during discussion

To understand the mechanisms that could underly the effects of both horizontal and top-down communication about transgender people, I here examine a number of intermediate outcomes, including attitudes, beliefs, norms, and behavior during the discussion (see Appendix J for more detail on the design of each measure). I show that the discussion effects are primarily mediated by a large shift in group-level norms of behavior towards transgender people (Table 4), i.e., changes in predictions about what other group members are doing. This shift comes along with very pro-trans behavior during the discussion itself. Later, I show the effects of the rights videos (Table A27, Table A26).

5.1 Effects of 3-person discussion

Norms. To examine whether descriptive norms of behavior towards transgender workers are affected by the treatment, I asked participants to make predictions about the *private* (outcome-round) hiring choices of others.

Beliefs about others' choices (community). First, participants 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 truthfully told that 20 other people in the study had been shown those pairs. They had to predict how many of those 20 picked each option. For the one pair that compared a male and a transgender, the predicted probability of selecting the transgender worker increased by a modest 4.3 p.p. (8.2%) in the discussion arm (Table 4, panel B, column 1).

Beliefs about others' choices (group). Participants then 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 the other person chose for two pairs of delivery options. The discussion caused a large increase in the predicted probability that a fellow group-member selects a transgender worker (Table 4, panel B, column 2; 23.9 p.p., $p<0.001$). The larger magnitude of the update in predictions *within* the group suggests that the effect of the discussion is likely to be mediated by a shift in group-level norms. In line with the claim that effects are mediated by group norms, controlling for group norms significantly attenuates the effect on discrimination, while controlling for community-wide norms does not (Table A25).

Attitudes. There were small or null effects of the discussion on measures of broader attitudes towards transgender people. The first measure of private attitudes I examine is a double list experiment (Droitcour et al., 2004; Glynn, 2013). It measured the proportion of people who agreed with the statement "In general, if I see a transgender person, I walk away." (see Appendix J.5 for more detail). Since the list experiment did not allow the enumerator or the researcher to infer *which* statements the participant agrees with from the list of statements, it preserved the anonymity of their responses and so was less likely to be vulnerable to social desirability concerns than standard self-reported attitude questions. The discussion

Table 4: *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***	0.239***
	(0.012) [<0.001]	(0.022) [<0.001]
Num. observations	2249	4465
Num. participants	2249	2238
Num. groups	751	751
Mean: No discussion (private)	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) [0.189]		
Anti-trans statement in list	0.204***		
	(0.033) [<0.001]		
3-person discussion		0.017**	
		(0.008) [0.040]	
Photo is trans × 3-person discussion			0.035
			(0.026) [0.185]
Photo is trans			−0.086***
			(0.025) [<0.001]
Num. observations	4498	4498	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Mean: No discussion (private)	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 J.9). 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 (Anderson, 2008).

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.

did not have a significant effect on the proportion of people who agreed with the sensitive anti-transgender statement in the list experiment (Table 4, panel A, column 1).

The second measure of attitudes was self-reported disapproval of discriminatory vignettes. Participants were presented with two scenarios depicting instances of discrimination against transgender individuals and were asked to evaluate whether the discriminator's actions were acceptable or wrong. There was a small improvement in attitudes (Table 4, panel A, column 2). The probability of saying that discrimination was wrong increased from an already high level in the control group (92.8%) to a slightly higher value (94.4%, p of difference: 0.04, effect size: 0.07 SD). Overall, the effect of the discussion does not seem to be driven by changes in very broad-based attitudes towards transgender individuals.

Beliefs about reliability. To measure whether there were changes in the perceived reliability of transgender workers as a result of the discussion, participants were asked to say how likely they think a certain worker was to complete a delivery if they were hired. Beliefs about the reliability of transgender workers were not significantly affected by the discussion (Table 4, panel A, column 3). While participants were 8.6 p.p. less likely to say that a worker is "likely" or "very likely" to complete the delivery when the worker is transgender (p <0.001), this does not vary significantly across treatment conditions (p=0.18).

5.2 *Effects of transgender rights videos*

While the discussion leads to changes in perceived norms and attitudes, the effects of the rights videos are mediated by changes in perceived norms and beliefs about reliability. The videos have a significant effect on perceived descriptive norms of discrimination (Table A26), in line with the *expressive law hypothesis* (Benabou & Tirole, 2011; Sunstein, 1996; McAdams & Rasmusen, 2004; Lane et al., 2019), which states that the law can affect behavior by signaling the prevailing social norm. After seeing either treatment video, participants predict that others will select will select transgender workers more, both in the wider community (2–3 p.p.), and in their group of 3 (4–6 p.p.). For community-wide norms, the effect of the legal rights video is similar to the effect of the group discussion (p=0.33). But for group-level norms, the discussion has a much stronger effect (p<0.001). This hints that the larger effects of the discussion may be mediated by the effects on group norms.

The videos also lead to small increases of around 6 p.p. (8%) in the probability that a participant reports that a transgender worker is likely to complete the delivery (Appendix Table A27, column 3). By contrast, neither video has a detectable effect on attitudes as measured by the list experiment or the questions on disapproval of discrimination (Appendix Table A27, columns 1-2).

5.3 *Behavior during the discussion*

To understand how the discussion encouraged people to discriminate less *after* the discussion, I here document evidence that participants exhibit very pro-trans behavior *during* the discussion.

Choices during the discussion (treatment round). The reduction in discrimination in the

private choices *after* the discussion is mirrored by large reductions in discrimination *during* the discussion. Table A29 shows that participants were 20 p.p. more likely to select a transgender worker in the collective choices during the discussion than the private choices made by those in the *No discussion (private)* condition ($p<0.001$). In these discussion choices, there was even *positive* discrimination in favor of transgender workers relative to non-transgender workers of around 11 p.p.[28] This suggests that participants persuade each other to discriminate less during the discussion, and that this spills over to later private choices.

Pro- and anti-trans statements in discussion. Participants communicated about transgender workers in a positive way. Enumerators observed the discussion and noted down (for each choice that involves a transgender worker) how many participants said something positive about the transgender worker, and how many said something negative about the transgender worker. Statements about transgender workers in the discussion were typically positive: participants were 5.7x more likely to say something positive about a transgender worker than to say something negative about them in the discussion (Figure A30).

Reasons cited in discussions. Figure A31 shows the different categories of reasons cited by participants in the discussion when they made their hiring choices, as measured by enumerator observations. Participants reacted differently when shown a choice-pair that included a transgender worker. They frequently stated explicitly that they made their choice because the worker was transgender, and were significantly more likely to cite pro-social rationales for their choices (difference: 32.5 p.p., $p<0.001$). For example, they were more likely to motivate their choice by saying they wanted to *give an opportunity* to the worker, or to *help* them. At the same time, they tended to underplay other factors such as items, worker details, or other characteristics. This shift towards pro-social reasoning is driven by groups who actually *chose* the transgender workers.

There is some correlative evidence that the shift towards pro-social reasoning may *persuade* others to discriminate less in the outcome round. First, the "listeners" in the 2-person discussion arm (who do not take part in the discussion themselves) were more likely to choose transgender workers in the outcome round when they heard pro-social and gender-based reasons in the discussion (Figure A32). Second, 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 more likely to cite pro-social reasons for their choices (Figure A33).

6 Mechanisms: What is behind the effect of the discussion?

In this section, I seek to understand how horizontal communication between privately discriminatory individuals can lead to strong reductions in discrimination, mediated by the emergence

[28] Anti-trans discrimination is lower on average in the control conditions in the earlier treatment round compared to the later outcome round, with transgender workers being around 9 p.p. less likely to be chosen. This could be rationalized by a self-signaling model, in which some participants try to prove to themselves that they are a "good person" by selecting a transgender worker when they first see one, but do not feel the need to do so in later rounds.

of a strong pro-trans norm. Here, I examine three candidate mechanisms that could explain this dynamic:

(1) *Correcting a misperceived norm*. Participants may initially overestimate how discriminatory their peers are. When they communicate, they realize that their peers are not as discriminatory as they thought, and so subsequently feel more comfortable selecting a transgender worker.

(2) *Virtue signaling*. Participants want to *appear* to be a "good person", i.e., not to be discriminatory, in a group setting. They therefore act positively towards transgender workers in the discussion and in doing so encourage others to discriminate less afterwards.

(3) *Persuasion and decision to speak up*. Participants change each others' preferences for selecting a transgender worker by sharing persuasive narratives. People who are more pro-trans are more vocal in discussions, leading groups to become overall less discriminatory.

Below, I document the evidence that channels (1) and (2) are not sufficient to explain the large effects of the discussion, whereas channel (3) is supported by the data and could explain large effects. I develop a model that examines channel (3) and describes the conditions under which the endogenous decision to speak up can generate large equilibrium reductions in discrimination.

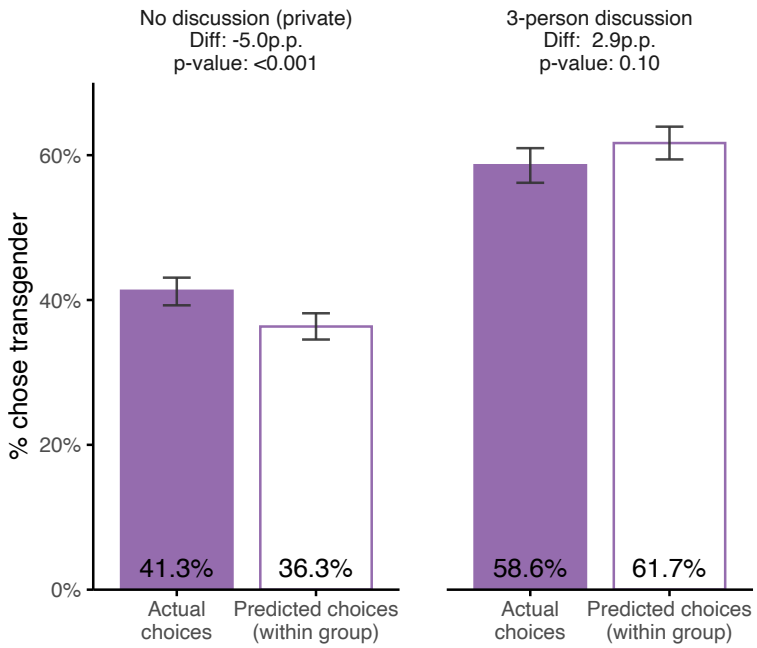
6.1 Correcting a misperceived norm

If participants initially overestimate how discriminatory their peers are, and if horizontal communication corrects that misperception, participants might feel more comfortable selecting a transgender worker after the discussion has finished.[29] Figure 5 examines this hypothesis by displaying participants' *within-group* predictions about others' private choices (described above in Section 5), and comparing them to the true probability of selecting a transgender worker. In line with this channel, control participants underestimate the probability that their group members select a transgender worker by 5.0 p.p. ($p<0.001$), suggesting an initial overestimation of discrimination.

However, a corrected misperception is not sufficient to explain the discussion's effects. While the discussion does stop participants overestimating discrimination, it also generates a large level-shift of roughly 20 p.p. in *both* the predictions and actual choices: people discriminate significantly less than would be the case if the control group's misperceptions were simply

[29]This idea is motivated by evidence in other contexts showing that correcting misperceptions about discriminatory norms can reduce anti-minority behavior (Bursztyn, González, & Yanagizawa-Drott, 2020). Alternatively, if participants initially *underestimated* how discriminatory their peers were, and this misperception were not corrected in the discussion, they may have faced face social pressure to discriminate less in the group discussion. Figure 5 shows that this does not fit the data.

Figure 5: *Evidence of misperceptions: predictions within a group of 3 (pairs involving transgender workers only)*



Notes: Sample includes all participants in the *3-person discussion* arm and the *No discussion (private)* arm, in both phases. 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).

corrected.[30] Since the control-group misperception was 5.0 p.p., and the total change in beliefs was 23.9 p.p., a simple back-of-the-envelope calculation would suggest that a perfectly precise correction of the misperception would account for only 21% (bootstrap 95% CI: [8.9%, 32.5%]) of the discussion's treatment effect. Thus, although correcting a misperceived norm might contribute to the discussion's impact, it is unlikely to account for the whole effect.

6.2 Virtue signaling

The virtue signaling channel proposes that participants have social image concerns, and so in group settings take pro-trans actions in order to not appear discriminatory (Bénabou & Tirole, 2006; DellaVigna et al., 2012; Bursztyn & Jensen, 2017). These pro-trans behaviors may persuade others to be less discriminatory after the discussion has ended.

Using the *No discussion (public)* arm, I test for virtue signaling by examining whether social

[30] A second piece of evidence against the misperception channel is based on *No discussion (public)* participants, who were told the *public* choices of others in their group before making predictions about *private* choices. They also had their misperceptions corrected (Figure A28, estimate of misperception: −1.8 p.p., p=0.17), but the effect on discrimination in this arm was much smaller than the effect of the discussion (Table 5).

image concerns alone can promote pro-trans choices in a group. Participants in this arm knew others would see their choices in the treatment round, but did not discuss those choices. If virtue signaling was driving behavior, we would therefore expect more pro-trans choices in this public setting.

Empirically, virtue signaling does not appear to be sufficient to explain the effects of the discussion. The *No discussion (public)* treatment did *not* make participants choose transgender workers more often in the treatment round on average (Table A34, p=0.46).[31] This was not because the treatment had no effect on behavior: participants within a group converged in their likelihood of selecting a transgender compared to the control group (p=0.06), suggesting that when choices were visible, participants tended to match the behavior of their group members (Appendix Table A35). There were also small or null effects on the *outcome* round (Table 5).[32]

6.3 Persuasion

A third channel that could explain the effects of the discussion and the emergence of a pro-trans norm is that (i) people persuade each other with the narratives and justifications they share during the discussion, and that (ii) persuasive communication is predominantly in favor of transgender workers, because the pro-trans participants are more vocal in the discussion.

6.3.1 Effect of listening to discussion

To test whether participants are persuaded by what they hear in the discussion, I examine the effects on the *listener* in the *2-person discussion arm*, who silently listens to other participants take part in a discussion. Listening leads to large and significant reductions in subsequent private discrimination (13.3 p.p., p<0.001, Table 5). This effect is not significantly different from the effect of speaking in either the 2-person discussion (p=0.86) or the 3-person discussion (p=0.21). Since the listener is silent, this suggests that the effect of the discussion is unlikely to operate through self-persuasion or self-consistency channels, where active participation in the discussion is crucial for generating reductions in discrimination (Falk & Zimmermann, 2017; Schwardmann et al., 2022). Instead, hearing the choices and justifications made by others in the discussion appears to be the key driver behind the treatment effects, so I interpret this as evidence for persuasive communication.[33]

[31] If pro-trans *communication* in the discussion is a costlier (stronger) virtue signal than pro-trans *choices*, the result here cannot rule out virtue signaling during discussions. However, this hypothesis seems unlikely, since selecting a transgender worker requires actually interacting with a trans-person, likely a higher cost for discriminators than non-binding pro-trans remarks.

[32] No significant effect is seen for *non-observers* ($p \in [0.14, 0.42]$), who were not told the public choices of others in their group before making their private outcome round choices. *Observers*, who were told the public choices of others in their group in advance, were around 5.4 p.p. more likely to select transgender workers in the private outcome round ($p \in [0.037, 0.096]$).

[33] The discussion reduces discrimination even on an outcome that is completely private (i.e., not observable by neighbors), adding to the evidence that participants are privately persuaded by what they hear (Section 7 and Table A40). I can also rule out that the listener effect is due to a correction of misperceptions, as per the arguments in Section 6.1.

Table 5: *Effect of phase 2 mechanism treatments 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 × 3-person discussion	0.190***	0.181***	
	(0.034) [<0.001]	(0.033) [<0.001]	
Worker is trans × Speaker (2-person discussion)	0.140***	0.127***	
	(0.032) [<0.001]	(0.031) [<0.001]	
Worker is trans × Listener (2-person discussion)	0.133***	0.133***	
	(0.040) [<0.001]	(0.039) [<0.001]	
Worker is trans × Observer (No discussion, public)	0.054**	0.048*	
	(0.029) [0.037]	(0.028) [0.055]	
Worker is trans × Non-observer (No discussion, public)	0.042	0.040	
	(0.036) [0.143]	(0.036) [0.135]	
3-person discussion	−0.003	0.005	0.176***
	(0.017) [0.854]	(0.016) [0.719]	(0.031) [<0.001]
Speaker (2-person discussion)	0.005	0.008	0.131***
	(0.016) [0.702]	(0.015) [0.538]	(0.028) [<0.001]
Listener (2-person discussion)	0.009	0.000	0.124***
	(0.020) [0.561]	(0.019) [0.971]	(0.033) [<0.001]
Observer (No discussion, public)	−0.003	−0.003	0.043*
	(0.015) [0.828]	(0.014) [0.825]	(0.026) [0.096]
Non-observer (No discussion, public)	−0.014	−0.010	0.021
	(0.020) [0.373]	(0.018) [0.470]	(0.031) [0.421]
Worker is trans	−0.208***		
	(0.018) [<0.001]		
Num. observations	13 308	13 308	4436
Num. participants	2218	2218	2218
Num. groups	741	741	741
Mean: no discussion (private), worker is non-trans	0.62	0.62	
Mean: no discussion (private), worker is trans	0.41	0.41	0.41
Controls		X	X
Controls interacted with worker is trans		X	
p(Observer=Non-observer)	0.744	0.816	0.489
p(Observer=Speaker)	0.015	0.018	0.004
p(Observer=Listener)	0.060	0.036	0.022
p(Observer=3-person discussion)	0.000	0.000	0.000
p(Speaker=Listener)	0.864	0.873	0.859
p(Speaker=3-person discussion)	0.204	0.161	0.208
p(Listener=3-person discussion)	0.214	0.292	0.204

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. The specification used is seen in equation 1, and is otherwise the same as Tables 1 and 2.

6.3.2 *Correlation between discussion behavior and discrimination*

Why did discriminatory participants persuade each other to be more *pro*-trans, rather than more *anti*-trans? I here document evidence that this was because the pro-trans individuals in the discussion were more vocal, and therefore persuaded other group members to be more pro-trans. I use the private choices *after* the discussion as a proxy for pro-trans behavior, and show that this is correlated with dominating the discussion of transgender workers (Table A36). Each additional transgender worker selected in the private outcome round is associated with being 11 p.p. (32%, p=0.03) more likely to speak first when faced with a choice involving a transgender worker, and 15 p.p. (27%, p=0.02) more likely to be rated by enumerators as dominating the discussion when faced with a transgender worker. This association is *specific* to choices that involve a transgender worker; these same participants are not more likely to dominate when faced with non-transgender choices. Since I did not collect baseline measures of discrimination (in order to reduce 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 persuasion is not *too* strong, initially anti-trans participants are unlikely to become more pro-trans than those who started off as pro-trans from the beginning. Moreover, the notion that pro-trans people are more likely to speak up is in line with the highly pro-trans pattern of communication documented in Section 5.3 (for example, statements about transgender workers were 5.7x more likely to say something positive than to say something negative).

6.4 *Model*

Motivated by the evidence that pro-trans participants are more vocal than anti-trans participants (when facing a choice that includes a transgender worker), I develop a model that derives the conditions under which a group of privately discriminatory individuals can persuade each other to discriminate less. More broadly, the model attempts to understand *why* horizontal communication can reduce discrimination.

The main result is that there is a "sweet spot" range of preferences which generates an equilibrium in which *only* pro-trans participants speak up in favor of transgender workers, and anti-trans participants stay silent. When baseline preferences are on average negative towards transgender workers (but not too negative), only pro-trans messages will be heard. This means that participants are on average *persuaded* to be more pro-trans.

The model's starting point is that participants care about fitting in with their group when making observable choices (Asch, 1956), motivated by the result that participants match their group member's behavior more in the *No discussion (public)* condition (Table A35). When no discussion is possible, the only way for a pro-trans group member to fit in with an anti-trans group is to discriminate. But when participants can *persuade* each other in a discussion, pro-trans people can also fit in with their group by persuading others to have pro-trans preferences. And because pro-trans people start off further from the existing discriminatory norm, they have a greater incentive to persuade others in their group. Pro-trans participants are therefore

more vocal, and persuade others to discriminate less, even after the discussion has ended.[34]

6.4.1 *Model set-up: During-discussion choices*

During the discussion, participants face a binary choice of whether to select a transgender worker ($Y_i = 1$) or not ($Y_i = 0$).[35] They have a private willingness to pay P_i for or against selecting a transgender worker, which is drawn from a uniform distribution with support $[\mu_P - R, \mu_P + R]$, with $R > 0$. This private preference is a reduced-form way of capturing both personal attitudes towards transgender people as well as personal norms about what the "right" thing to do is in a hiring choice. Empirically, there is a large willingness to pay in to avoid transgender workers in the control group, so I assume that at baseline participants are on average anti-transgender, i.e., $\mu_P < 0$.

Participants can incur a cost c to send a persuasive message to their group members. This can be a pro-trans message ($S_i = 1$), an anti-trans message ($S_i = -1$), or no message ($S_i = 0$). After hearing a pro-trans message from someone else in their group, a participant's private preference will be increased by α, where α denotes the change in willingness to pay, i.e., the "persuasiveness" of the messages. Similarly, hearing an anti-trans message will lead to a *decrease* of post-discussion preference by α.[36] So i's post-discussion preferences will be $\tilde{P}_i = P_i + \alpha(S_j + S_k)$, where S_j and S_k are the messages sent by the other two members of i's group.

The timing of the game is as follows. First, participants observe their own private preference P_i, then all three participants in a group simultaneously choose *both* Y_i and S_i. Everyone observes the full set of actions $(Y_i, S_i)_{i=1,2,3}$ and then form expectations about the mean post-discussion preferences of others $\tilde{P}_{-i} := (\tilde{P}_j + \tilde{P}_k)/2$, given the full set of actions taken and the persuasion that results from those actions.

A strategy $\sigma_i : [\mu_P - R, \mu_P + R] \to \{0, 1\} \times \{-1, 0, 1\}$ is a mapping from private preference P_i to an action (Y_i, S_i). Focusing on pure strategies, a participant with preference P_i will choose

[34] The model does not require that pro-trans participants have a *stronger* preference than anti-trans preference. Empirically, the distribution of preferences is not skewed (Figure A37), so this cannot be the explanation for my results.

[35] I abstract away from two empirical features of the discussion. First, I do not model the collective decision-making process during the discussion, instead considering that participants choose a worker Y_i directly. Second, I do not model the dynamics of the discussion. I do not account for these features because (i) it makes the analysis tractable, and (ii) I do not have data on the dynamic process of the model that would allow me to validate the predictions of a dynamic model.

[36] I assume that pro-trans messages are as persuasive as anti-trans messages. An alternative model with *asymmetric persuasion*, where pro-trans messages are inherently more persuasive than anti-trans messages, could generate similar conclusions to the model I outline.

an action to maximize expected utility, taking as given the strategies of others σ_{-i}:

$$\max_{Y_i,S_i} \mathbb{E}_i[U_i(Y_i,S_i|P_i,\sigma_{-i}))] = V(Y_i) + P_i \cdot \mathbb{1}\{Y_i = 1\} + \gamma_0 \, \mathbb{E}_{-i}[\tilde{P}_i|Y_i,S_i,\sigma_{-i}] \\ - \gamma_1 \left(\mathbb{E}_{-i}[\tilde{P}_i|Y_i,S_i,\sigma_{-i}] - \mathbb{E}_i[\tilde{P}_{-i}|S_i,\sigma_{-i}] \right)^2 \\ - c \cdot \mathbb{1}\{S_i \in \{-1,1\}\} \tag{2}$$

$V(Y_i) \in \mathbb{R}$ is the value of the items offered by the option Y_i, and P_i is incurred as a cost or benefit only when selecting the transgender worker. When making choices that are visible to their group, participants have two types of social image concerns. First, they have a preference $\gamma_0 \in [0,\infty)$ to *virtue signal* that they are a "good person", i.e. that they have a high $\tilde{P}_i$. Second, they have a preference $\gamma_1 \in [0,\infty)$ to *conform* to their group, i.e., they do not wish to be perceived to have a $\tilde{P}_i$ that deviates too far from their group's.[37] The expectations in the utility function use post-discussion preferences $\tilde{P}_i$, so they take into account a prediction of how much persuasion will occur in equilibrium (and therefore depend on α). I assume the full set of parameters $(\alpha, \mu_P, R, c, \gamma_0, \gamma_1)$ are the same for all i and are common knowledge, whereas P_i is only known to i.

6.4.2 *Equilibria*

I focus on symmetric Bayesian Nash equilibria in pure strategies. I restrict to equilibria in which all participants who choose the same Y_i also choose the same S_i (i.e., there is a one-to-one mapping between Y_i and S_i), which I call *homogeneous* equilibria. This effectively simplifies the action space so that equilibrium choices are characterized by a single threshold (as in Bénabou & Tirole, 2006). In other words, there is a threshold where the marginal agent has $P_i = P^*$, such that all agents with $P_i > P^*$ choose $Y_i = 1$, and all agents with $P_i \leq P^*$ choose $Y_i = 0$. P^* is always unique when P_i is drawn from a uniform distribution. Under assumptions about communication being sufficiently cheap, not too persuasive, and conformity pressures not being too large, there are three types of equilibria.

Proposition 1. *Assume $c < c^* = \alpha\gamma_1 R - \alpha^2\gamma_1$, $\alpha < R/2$, and $\gamma_1 < \frac{R}{2R^2-\alpha R-\alpha^2}$. Then the following three equilibria (denoted $Q = \{SN, SS, NS\}$) exist:*

1. *If $\mu_P \in [\eta^{SN}_{Y=0}, \kappa^{SN}_{Y=0,S\neq-1}]$, i.e., when baseline preferences are anti-trans, a "Send-NoSend" (SN) equilibrium exists, in which only participants who choose $Y_i = 1$ send a message $S_i = 1$, and others do not send a message ($S_i = 0$).*

2. *If $\mu_P \in [\kappa^{SS}_{Y=0,S\neq0}, \kappa^{SS}_{Y=1,S\neq0}]$, i.e., when baseline attitudes are mid-ranged, a "Send-Send" (SS) equilibrium exists, in which participants who choose $Y_i = 1$ send $S_i = 1$, and participants who choose $Y_i = 0$ send $S_i = -1$.*

3. *If $\mu_P \in [\kappa^{NS}_{Y=1,S\neq1}, \eta^{NS}_{Y=1}]$, i.e., when baseline attitudes are pro-trans, a "NoSend-Send" (NS)*

[37] I assume convex costs of deviating from one's group. This means that the marginal benefit of persuading others will be *greater* when their action would imply that they are further away from the group norm. The virtue signaling and conformity motives can cancel out when persuasion is not possible (see Appendix F.2), explaining the null result on the *No discussion (public)* treatment in the data.

equilibrium exists, in which only participants who choose $Y_i = 0$ send a message $S_i = -1$, and others do not send a message ($S_i = 0$).

Proofs are in Appendix F. This proposition implies that there is a sweet spot range of baseline preferences $\mu_P \in [\eta^{SN}_{Y=0}, \kappa^{SN}_{Y=0,S\neq-1}]$ which permit an equilibrium in which *only* pro-trans people send persuasive messages in the discussion. Figure A38 illustrates an example of the parameter space where each equilibrium exists.[38]

To gain intuition on the result, consider that the incentive to send a message is greatest when a participant's action Y_i takes them far away from the baseline norm. This is because the message allows them to change others' private preferences to be more like them, decreasing the cost of deviating from the group norm. Because of this, when baseline preferences are anti-trans, only those who choose transgender workers are willing to incur the cost to send a message. The incentive of those who *don't* choose a transgender worker is weak; they already conform to the group norm. This is why the "Send-NoSend" equilibrium exists when μ_P is low. When μ_P is too low, however, i.e., $\mu_P < \eta^{SN}_{Y=0}$, preferences are so anti-trans that everyone chooses $Y_i = 0$: no-one selects a transgender worker, and no-one sends a pro-trans message. This provides a logic for why, when baseline preferences are discriminatory (but not too discriminatory), we empirically observe *more* pro-trans than anti-trans communication.

6.4.3 *Post-discussion choices*

Here I examine how post-discussion choices are affected by the discussion. These choices are (i) private, and so not subject to social image concerns, and (ii) based on participants' post-discussion private preferences $\tilde{P}_i$, which have changed based on messages received during the discussion. Participants will therefore maximize:

$$\max_{\tilde{Y}_i} E[\tilde{U}_i(\tilde{Y}_i)] = \quad V(\tilde{Y}_i) + \tilde{P}_i = V(\tilde{Y}_i) + P_i + \alpha(S_j + S_k)$$

where $\tilde{Y}_i$ is i's new post-discussion choice of worker, and S_j and S_k are the messages sent by the others during the discussion.[39] If baseline preferences μ_P are in the sweet spot range, post-discussion choices can be less discriminatory because a group can be in a "Send-NoSend" equilibrium, in which only pro-trans people send pro-trans messages, so $S_j + S_k > 0$. Participants are on net *persuaded* to discriminate less after the discussion. By contrast, participants in the control group (who do not communicate with each other) will continue to

[38] Under the assumptions in Proposition 1, $\eta^{SN}_{Y=0} < \kappa^{SN}_{Y=0,S\neq-1} < \kappa^{SS}_{Y=0,S\neq0} < \kappa^{SS}_{Y=1,S\neq0} \gtreqless \kappa^{NS}_{Y=1,S\neq1} < \eta^{NS}_{Y=1}$. The strict inequalities imply there are some gaps in the parameter space where homogeneous equilibria do not exist. And since $\kappa^{SS}_{Y=1,S\neq0} \gtreqless \kappa^{NS}_{Y=1,S\neq1}$, there are sometimes ranges where both NS and SS exist as equilibria.

[39] I assume that the pre- and post-discussion choices are separable in the sense that participants do not need to take their post-discussion choices into account when considering what to choose in the discussion. This holds if I assume risk neutrality and since Y_i is independent from $\tilde{Y}_i$.

have the same discriminatory preferences.[40] The value of μ_P that generates the most pro-trans persuasion will be just below the upper limit of the region where SN exists (i.e., just below $\kappa^{SN}_{Y=0,S\neq-1}$), since this is the value of μ_P that maximizes the number of people sending pro-trans messages *without* inducing anti-trans participants to also start persuading (see Figure A39).

In summary, under reasonable assumptions, there is a sweet-spot range of average baseline preferences in which only pro-trans people send pro-trans messages, and anti-trans participants do not send a message. If people are sufficiently anti-trans, anti-trans participants will not want to send a persuasive message because the marginal benefit of doing so for them will be small: they already match the group norm, so have little incentive to persuade others to match their preferences. By contrast, the pro-trans people are far from the discriminatory group norm, and so have more incentive to try and persuade others to match their preferences. In this sweet-spot range, only pro-trans persuasive messages are circulating in the discussion. This explains why people are on average significantly more pro-trans after the discussion has ended.

7 Alternative mechanisms

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

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 evaluate whether the treatment effects remained when shutting down this channel, I use a series of supplementary hiring choices. These *private grocery pick-up* choices (described in more detail in Appendix J.7) were designed to be more robustly private than the main outcome in two ways. First, so that neighbors would not know which worker was chosen, the participant had to pick up grocery items from the team office, instead of receiving the delivery at home. Second, I adapted 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 (Table A40).[41] The discussion treatment effect on this outcome is large, although slightly

[40] I also show that a "NoSend-NoSend" equilibrium exists when the cost of communication c is sufficiently high (Appendix F). This provides one explanation for why participants have not *already* communicated about transgender workers in equilibrium before the experiment takes place. For example, if participants are rarely in situations in which one is required to talk about transgender persons, the opportunity cost of talking about them (instead of other subjects) may be high. Prompting discussion can therefore be conceived of as exogenously decreasing c.

[41] Discrimination in the *No discussion (private)* arm was stronger for these private outcomes than for the main hiring elicitation (29.4 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.

smaller in magnitude than the main hiring outcome (12.5 p.p., $p<0.001$). The *legal rights* video also reduces discrimination significantly, with a similar magnitude to the main outcome (Table A41). 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 out that such concerns play some role.

Deliberation. 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 (Devine, 1989; Devine & Monteith, 1993; Plant & Devine, 2009; Devine et al., 2012). There is some evidence for such increased deliberation. Discussion participants take on average 2.2 seconds (27%) longer in the individual outcome-round choices (p <0.001, Figure A42, panel B), and are *less* likely to select a dominated option in the outcome round if they have been in a group discussion (p=0.02, Table A15, column 1), suggesting they are being more attentive. However, it remains unclear how much this drives the treatment effects on the *outcome round* choices, since in the outcome round, longer response times are not correlated with being more likely to select a transgender worker (p=0.43, Table A43, column 3).

Other photo characteristics. If the transgender worker photos were observably different to 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. 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. (Table A44). 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.

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, 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 (Table A45, column 1), and the effect on discrimination is not significantly stronger for participants who remembered the word transgender (Table A46, column 4).

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 (de Quidt et al., 2018). 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). 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 (Table A45, columns 2 and 3; Figure A48), and there is no detectable difference in the treatment effects for people that do and do not correctly guess the study purpose (Table A46, columns 1 and 2). While the rights videos did increase the likelihood of a participant correctly guessing the purpose of the experiment (Table A47), those who correctly guessed did not drive the reductions in discrimination seen in the discussion groups (Table A49, columns 1 and 2).

Social desirability bias. To measure a participant's propensity to give socially desirable answers, at baseline I elicited a shortened version of the Crowne & Marlowe (1960) module, which has been used elsewhere in India for a similar purpose (Dhar et al., 2022). The questions ask whether the respondent has a number of "too good to be true" traits (see Appendix J.3). 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 (Table A46, column 3; Table A49, column 3).

Increased stakes. 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 (Table A50, 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. The evidence on the rights videos is more mixed: the point estimate suggests that the effect of the videos is smaller when participants are offered 3 deliveries, although I cannot detect a significant difference (Table A51). The effects of informing people about transgender rights may therefore be attenuated in higher-stakes situations.

Other mechanisms behind the discussion. A number of other features of the process differ

between participants involved in a discussion and those that select individually.[42] However, these features cannot explain the entire effects of the discussions, because they are also shared by the *non-observers* in the *No discussion (public)* arm, who are not significantly more likely to select a transgender worker in the outcome round.

8 Conclusion

Involving majority-group members in a group discussion and hiring decision can sharply reduce discrimination against transgender people in a real-stakes hiring decision. Even though the discussion I evaluate lasts only 10 minutes, it also has impacts on medium-run choices. My results act as a proof-of-concept that the horizontal communication that naturally arises when discussing a minority can lead to large reductions in discrimination. I also show that top-down communication about the legal rights of a minority can significantly reduce discrimination in the short-run, although the effects are substantially smaller.

A key remaining uncertainty is the extent to which these results generalize to other contexts and other minorities. An important limitation of the study is the focus on the transgender community in India, and the concern that the specific social dynamics driving behavior towards that community do not generalize to other minorities. There are therefore several important research avenues that are important for understanding whether the mechanisms I examine are present in other contexts.

First, does pro-minority horizontal communication arise endogenously in other group settings? Other research suggests that virtue signaling motives can discourage people from expressing anti-minority views (Bursztyn et al., 2023; Bursztyn, Egorov, & Fiorin, 2020; Braghieri, 2021), providing one mechanism that may be common to other contexts. I show evidence that horizontal communication can also be beneficial when pro-minority individuals speak up and persuade others to discriminate less, which can occur when attitudes are in a sweet spot – discriminatory, but not too discriminatory. Despite economically important anti-transgender discrimination in the control group, this discrimination may be lightly held, and borne of unfamiliarity rather than deep animosity. This suggests that discussions will be most effective when there are many people on the margin of not discriminating, and when there are some people willing to advocate for a discriminated group.

Such scenarios may not be uncommon. For example, research on other stigmatized groups — such as those experiencing homelessness, poverty, or disability — has shown that sympathy, pity, and guilt can motivate supportive actions, even while coexisting with stigmatizing attitudes (e.g., O'Driscoll & Feather, 1985; Iyer et al., 2003; Mallett et al., 2008; Harth et al., 2008; Thomas et al., 2009; Tsai et al., 2017; Lantos et al., 2020; Dull et al., 2021). Conversely, horizontal communication is likely to be less effective when anti-minority attitudes are very deep-set, or when people are being asked to engage in pro-minority actions that are significantly more

[42] These differences include: (i) the group setting, which may affect participants' mood or pro-sociality; (ii) the icebreaker discussion, which may relax them or make them less suspicious; (iii) being shown the worker profiles on paper sheets instead of on the enumerator's tablet; and (iv) the longer delay between the treatment round and the outcome round, due to having to move from a common group space to a private space.

costly. For example, while participants changed their willingness to interact with a transgender worker for 15 minutes, more intensive interventions would be needed to change participants' willingness to have a transgender neighbor, work with a transgender colleague for a year, become friends with a transgender person.

Second, my results raise the question of why the horizontal communication that reduces discrimination has not already occurred in equilibrium. In the follow-up survey, only 31% of the control group had talked about transgender people in the time since the main survey (even after having gone through the survey itself, which clearly involved transgender people). One possibility is that people are exploiting "moral wiggle room".[43] They avoid talking about transgender people in order to avoid having to act pro-socially towards them; they would prefer to act selfishly towards them without making it explicit that they are discriminating. Alternatively, discrimination could come hand-in-hand with a lack of social contact between transgender and non-transgender people, meaning that transgenderism is rarely raised as a topic. Either of these reasons could explain why creating a situation in which transgender workers are explicitly discussed can have large effects on discrimination.

It will also be important to build policies based on the insight that group communication can reduce discrimination under the right circumstances. First, we should design and evaluate policies that create discussions at scale to change attitudes towards gender minorities. Previous work shows that it may be possible to change discriminatory attitudes by running interventions in schools (Dhar et al., 2022), or by door-to-door canvassing (Kalla & Broockman, 2020; Broockman & Kalla, 2016). My results raise the possibility of reducing discrimination without even having to *lead* a discussion; instead, just creating a scenario where minorities are naturally discussed at all may be sufficient in some contexts. One important caveat is that the short 10-minute discussion in my study only generates small medium-term impacts on discrimination. Policies likely require more intensive and repeated interventions to have larger and longer-run effects.

Second, my results suggest that under the right conditions, groups discriminate much less than individuals. This implies that in high-stakes decisions where discrimination might take place (in hiring, housing, college admissions, etc.), it is especially important to design a decision environment that is conducive to egalitarian decision-making. In particular, we should investigate further how group dynamics and collective hiring choices may affect discrimination, building on existing work that has examined the effect of different compositions of hiring and academic selection committees (e.g., M. F. Bagues & Esteve-Volart, 2010; M. Bagues et al., 2017). This requires going beyond the conventional economic perspective, which views discrimination primarily through the lens of individual decision-making.

[43] See, e.g., Dana et al. (2007); Lazear et al. (2012); Hamman et al. (2010); Dana et al. (2006); Andreoni et al. (2017) for examples of this moral wiggle room effect.

References

Abbate, N., Berniell, I., Coleff, J., Laguinge, L., Machelett, M., Marchionni, M., ... Pinto, M. F. (2023). Discrimination Against Gay and Transgender People in Latin America: A Correspondence Study in the Rental Housing Market.

Agoramoorthy, G., & Hsu, M. J. (2015, August). Living on the Societal Edge: India's Transgender Realities. *Journal of Religion and Health*, *54*(4), 1451–1459. doi: 10.1007/s10943-014-9987-z

Aigner, D. J., & Cain, G. G. (1977, January). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, *30*(2), 175–187. doi: 10.1177/001979397703000204

Aksoy, C. G., Carpenter, C. S., De Haas, R., Dolls, M., & Windsteiger, L. (2021). Reducing Sexual-Orientation Discrimination: Experimental Evidence from Basic Information Treatments. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3995522

Aksoy, C. G., Carpenter, C. S., De Haas, R., & Tran, K. D. (2020, May). Do laws shape attitudes? Evidence from same-sex relationship recognition policies in Europe. *European Economic Review*, *124*, 103399. doi: 10.1016/j.euroecorev.2020.103399

Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.

Ambrus, A., Greiner, B., & Pathak, P. A. (2015, September). How individual preferences are aggregated in groups: An experimental study. *Journal of Public Economics*, *129*, 1–13. doi: 10.1016/j.jpubeco.2015.05.008

Anderson, M. L. (2008, December). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. doi: 10.1198/016214508000000841

Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2021, April). Predicting social tipping and norm change in controlled experiments. *Proceedings of the National Academy of Sciences*, *118*(16), e2014893118. doi: 10.1073/pnas.2014893118

Andreoni, J., Rao, J. M., & Trachtman, H. (2017). Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. *journal of political economy*, 29.

Andrew, A., Krutikova, S., Smarrelli, G., & Verma, H. (2022, September). *Gender norms, violence and adolescent girls' trajectories: Evidence from a field experiment in India* (Tech. Rep.). The IFS. doi: 10.1920/wp.ifs.2022.4122

Angerer, S., Waibel, C., & Stummer, H. (2018, January). *Discrimination in Health Care: A Field Experiment on the Impact of Patients' Socio-Economic Status on Access to Care* (SSRN Scholarly Paper No. 3036000). Rochester, NY. doi: 10.2139/ssrn.3036000

Arrow, K. J. (1973, July). Higher education as a filter. *Journal of Public Economics*, *2*(3), 193–216. doi: 10.1016/0047-2727(73)90013-3

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. doi: 10.1037/h0093718

Baba, R., & Sogani, R. (2018, June). Transgender Health and Healthcare in India: A Review..

Badgett, M. V. L. (2014). The economic cost of stigma and the exclusion of LGBT people: A case study of India.

Badgett, M. V. L., Carpenter, C. S., & Sansone, D. (2021, May). LGBTQ Economics. *Journal of Economic Perspectives*, *35*(2), 141–170. doi: 10.1257/jep.35.2.141

Badgett, M. V. L., Waaldijk, K., & Rodgers, Y. v. d. M. (2019, August). The relationship between LGBT inclusion and economic development: Macro-level evidence. *World Development*, *120*, 1–14. doi: 10.1016/j.worlddev.2019.03.011

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017, April). Does the Gender Composition of Scientific Committees Matter? *American Economic Review*, *107*(4), 1207–1238. doi: 10.1257/aer.20151211

Bagues, M. F., & Esteve-Volart, B. (2010, October). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *Review of Economic Studies*, *77*(4), 1301–1328. doi: 10.1111/j.1467-937X.2009.00601.x

Banerjee, A., La Ferrara, E., & Orozco-Olvera, V. H. (2019). The Entertaining Way to Behavioral Change: Fighting HIV with MTV. *NBER Working Paper*(26096).

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009, November). Powerful Women: Does Exposure Reduce Bias?*. *The Quarterly Journal of Economics*, *124*(4), 1497–1540. doi: 10.1162/qjec.2009.124.4.1497

Becker, G. S. (1957). *The Economics of Discrimination* (d. edition, Ed.). Chicago, IL: University of Chicago Press.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014, April). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, *81*(2), 608–650. doi: 10.1093/restud/rdt044

Bénabou, R., Falk, A., & Tirole, J. (2020). Narratives, Imperatives, and Moral Persuasion. , 56.

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *The American Economic Review*, *96*(5), 35.

Benabou, R., & Tirole, J. (2011, November). *Laws and Norms* (Tech. Rep. No. w17579). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w17579

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227–1274. doi: 10.1037/bul0000067

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2023, September). Inaccurate Statistical Discrimination: An Identification Problem. *The Review of Economics and Statistics*, 1–45. doi: 10.1162/rest_a_01367

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., & Eccles, J. (2006, December). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, *96*(5), 1890–1905. doi: 10.1257/aer.96.5.1890

Braghieri, L. (2021). Political Correctness, Social Image, and Information Transmission. doi: 10.1257/rct.5063-1.1

Broockman, D., & Kalla, J. (2016, April). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224. doi: 10.1126/science.aad9713

Burn, I. (2020, May). The Relationship between Prejudice and Wage Penalties for Gay Men in the United States. *ILR Review*, *73*(3), 650–675. doi: 10.1177/0019793919864891

Bursztyn, L., Egorov, G., & Fiorin, S. (2020, November). From Extreme to Mainstream: The Erosion of Social Norms. *American Economic Review*, *110*(11), 3522–3548. doi: 10.1257/aer.20171175

Bursztyn, L., Egorov, G., Haaland, I., Rao, A., & Roth, C. (2023, January). Justifying Dissent. *The Quarterly Journal of Economics*, qjad007. doi: 10.1093/qje/qjad007

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020, October). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review*, *110*(10), 2997–3029. doi: 10.1257/aer.20180975

Bursztyn, L., & Jensen, R. (2017). Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure. *Annual Review of Economics*, *9*(1), 131–153. doi: 10.1146/annurev-economics-063016-103625

Button, P., Dils, E., Harrell, B., Fumarco, L., & Schwegman, D. (2020, December). *Gender Identity, Race, and Ethnicity Discrimination in Access to Mental Health Care: Preliminary Evidence from a Multi-Wave Audit Field Experiment* (Tech. Rep. No. w28164). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w28164

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995, September). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. doi: 10.1137/0916069

Carpenter, C. S., Eppink, S. T., & Gonzales, G. (2020, May). Transgender Status, Gender Identity, and Socioeconomic Outcomes in the United States. *ILR Review*, *73*(3), 573–599. doi: 10.1177/0019793920902776

Center, N. O. R. (2014). *General Social Survey.* Harvard Dataverse. doi: 10.7910/DVN/26571

Chakrapani, Babu, P., & Ebenezer, T. (2004, June). Hijras in sex work face discrimination in the Indian health-care system..

Chakrapani, V., Newman, P. A., Shunmugam, M., & Dubrow, R. (2011, December). Barriers to free antiretroviral treatment access among kothi-identified men who have sex with men and aravanis (transgender women) in Chennai, India. *AIDS Care*, *23*(12), 1687–1694. doi: 10.1080/09540121.2011.582076

Charles, K. K., & Guryan, J. (2008, October). Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination. *Journal of Political Economy*, *116*(5), 773–809. doi: 10.1086/593073

Chen, D. L., & Yeh, S. (2014, August). The construction of morals. *Journal of Economic Behavior & Organization*, *104*, 84–105. doi: 10.1016/j.jebo.2013.10.013

Chingwete, A., Richmond, S., & Alpin, C. (2014). Support for African women's equality rises. *Afrobarometer Policy Paper #8*.

Christensen, P., & Timmins, C. (2023, November). The Damages and Distortions from Discrimination in the Rental Housing Market*. *The Quarterly Journal of Economics*, *138*(4), 2505–2557. doi: 10.1093/qje/qjad029

Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021, January). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, *148*, 102582. doi: 10.1016/j.jdeveco.2020.102582

Corno, L., La Ferrara, E., & Burns, J. (2022, December). Interaction, Stereotypes, and Performance: Evidence from South Africa. *American Economic Review*, *112*(12), 3848–3875. doi: 10.1257/aer.20181805

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002, March). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378.

Crowne, D. P., & Marlowe, D. (1960). Marlowe-Crowne social desirability scale. *Journal of Consulting Psychology*.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201. doi: 10.1016/j.obhdp.2005.10.001

Dana, J., Weber, R. A., & Kuang, J. X. (2007, July). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80. doi: 10.1007/s00199-006-0153-z

Davis, J. H. (1973, March). Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, *80*(2), 97–125. doi: 10.1037/h0033951

de Quidt, J., Haushofer, J., & Roth, C. (2018, November). Measuring and Bounding Experimenter Demand. *American Economic Review*, *108*(11), 3266–3302. doi: 10.1257/aer.20171330

Delhi, U. N. (2018). *Experiences of bullying in schools: A survey among sexual/gender minority youth in Tamil Nadu* (Tech. Rep.).

DellaVigna, S., List, J. A., & Malmendier, U. (2012, February). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*(1), 1–56. doi: 10.1093/qje/qjr050

Devine, P. (1989, January). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. doi: 10.1037/0022-3514.56.1.5

Devine, P., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012, November). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. doi: 10.1016/j.jesp.2012.06.003

Devine, P., & Monteith, M. (1993, December). The Role of Discrepancy-Associated Affect in Prejudice Reduction.
doi: 10.1016/B978-0-08-088579-7.50018-1

Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. , 66.

Dixit, V., Garg, B., Mehta, N., Kaur, H., & Malhotra, R. (2023, April). The Third Gender in a Third World Country: Major Concerns and the "AIIMS Initiative". *Journal of Human Rights and Social Work*, 1–6. doi: 10.1007/s41134-023-00238-3

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (2004). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185–210.

Drydakis, N. (2022, April). Sexual orientation and earnings: A meta-analysis 2012–2020. *Journal of Population Economics*, *35*(2), 409–440. doi: 10.1007/s00148-021-00862-1

Dull, B. D., Hoyt, L. T., Grzanka, P. R., & Zeiders, K. H. (2021, June). Can White Guilt Motivate Action? The Role of Civic Beliefs. *Journal of Youth and Adolescence*, *50*(6), 1081–1097. doi: 10.1007/s10964-021-01401-7

Falk, A., & Zimmermann, F. (2017, July). Consistency as a Signal of Skills. *Management Science*, *63*(7), 2197–2210. doi: 10.1287/mnsc.2016.2459

Fernández, R. (2013, February). Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *American Economic Review*, *103*(1), 472–500. doi: 10.1257/aer.103.1.472

Flores, A. R. (2015, July). Attitudes toward transgender rights: Perceived knowledge and secondary interpersonal contact. *Politics, Groups, and Identities*, *3*(3), 398–416. doi: 10.1080/21565503.2015.1050414

Folke, O., & Rickne, J. (2022, November). Sexual Harassment and Gender Inequality in the Labor Market*. *The Quarterly Journal of Economics*, *137*(4), 2163–2212. doi: 10.1093/qje/qjac018

Funk, P. (2007, April). Is There An Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines. *American Law and Economics Review*, *9*(1), 135–159. doi: 10.1093/aler/ahm002

Galbiati, R., Henry, E., Jacquemet, N., & Lobeck, M. (2020). How Laws Affect the Perception of Norms: Empirical Evidence from the Lockdown. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3684710

Ganju, D., & Saggurti, N. (2017, August). Stigma, violence and HIV vulnerability among transgender persons in sex work in Maharashtra, India. *Culture, Health & Sexuality*, *19*(8), 903–917. doi: 10.1080/13691058.2016.1271141

Glover, D., Pallais, A., & Pariente, W. (2017, August). Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores*. *The Quarterly Journal of Economics*, *132*(3), 1219–1260. doi: 10.1093/qje/qjx006

Glynn, A. N. (2013, January). What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly*, *77*(S1), 159–172. doi: 10.1093/poq/nfs070

Golman, R. (2022, July). *Acceptable Discourse: Social Norms of Beliefs and Opinions* (SSRN Scholarly Paper No. 4160955). Rochester, NY. doi: 10.2139/ssrn.4160955

Gómez, Á., Tropp, L. R., Vázquez, A., Voci, A., & Hewstone, M. (2018, May). Depersonalized extended contact and injunctive norms about cross-group friendship impact intergroup orientations. *Journal of Experimental Social Psychology*, *76*, 356–370. doi: 10.1016/j.jesp.2018.02.010

Granberg, M., Andersson, P. A., & Ahmed, A. (2020, August). Hiring Discrimination Against Transgender People: Evidence from a Field Experiment. *Labour Economics*, *65*, 101860. doi: 10.1016/j.labeco.2020.101860

Gulesci, S., Lombardi, M., & Ramos, A. (2023). Telenovelas and Attitudes toward the LGBTIQ Community in Latin America.

Gulesci, S., Smerdon, D., Jindani, S., Sulaiman, M., Ferrara, E. L., & Young, H. P. (2021). A Stepping Stone Approach to Understanding Harmful Norms. , 49.

Hamman, J. R., Loewenstein, G., & Weber, R. A. (2010, September). Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship. *American Economic Review*, *100*(4), 1826–1846. doi: 10.1257/aer.100.4.1826

Harth, N. S., Kessler, T., & Leach, C. W. (2008, January). Advantaged group's emotional

reactions to intergroup inequality: The dynamics of pride, guilt, and sympathy. *Personality & Social Psychology Bulletin*, *34*(1), 115–129. doi: 10.1177/0146167207309193

Hedegaard, M. S., & Tyran, J.-R. (2018, January). The Price of Prejudice. *American Economic Journal: Applied Economics*, *10*(1), 40–63. doi: 10.1257/app.20150241

Hjort, J. (2014, November). Ethnic Divisions and Production in Firms*. *The Quarterly Journal of Economics*, *129*(4), 1899–1946. doi: 10.1093/qje/qju028

IPSOS. (2018, January). *Global Attitudes Toward Transgender People.* https://www.ipsos.com/en/global-attitudes-toward-transgender-people.

Iyer, A., Leach, C. W., & Crosby, F. J. (2003, January). White guilt and racial compensation: The benefits and limits of self-focus. *Personality & Social Psychology Bulletin*, *29*(1), 117–129. doi: 10.1177/0146167202238377

Jayachandran, S. (2020). Social Norms as a Barrier to Women's Employment in Developing Countries. , 28.

John, A., & Orkin, K. (2022, June). Can Simple Psychological Interventions Increase Preventive Health Investment? *Journal of the European Economic Association*, *20*(3), 1001–1047. doi: 10.1093/jeea/jvab052

Kalla, J. L., & Broockman, D. E. (2020, May). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review*, *114*(2), 410–425. doi: 10.1017/S0003055419000923

Kalra, G. (2012, August). Hijras: The unique transgender culture of India. *International Journal of Culture and Mental Health*, *5*(2), 121–126. doi: 10.1080/17542863.2011.570915

Kerala Development Society. (2017). Study on human rights of transgender as a third gender. *The National Human Rights Commission*.

Kerr, N. L. (1981, October). Social transition schemes: Charting the group's road to agreement. *Journal of Personality and Social Psychology*, *41*(4), 684–702. doi: 10.1037/0022-3514.41.4.684

Kerr, N. L., Stasser, G., & Davis, J. H. (1979). Model testing, model fitting, and social decision schemes. *Organizational Behavior & Human Performance*, *23*, 399–410. doi: 10.1016/0030-5073(79)90006-0

Klawitter, M. (2015, January). Meta-Analysis of the Effects of Sexual Orientation on Earnings. *Industrial Relations: A Journal of Economy and Society*, *54*(1), 4–32. doi: 10.1111/irel.12075

Kumar, G., Suguna, A., Suryawanshi, D. M., Surekha, A., Rajaseharan, D., & Gunasekaran, K. (2022, November). Exploring the discrimination and stigma faced by transgender in Chennai city–A community-based qualitative study. *Journal of Family Medicine and Primary Care*, *11*(11), 7060–7063. doi: 10.4103/jfmpc.jfmpc_1037_22

Kuran, T. (1987, September). Preference Falsification, Policy Continuity and Collective Conservatism. *The Economic Journal*, *97*(387), 642. doi: 10.2307/2232928

Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World politics*, *44*(1), 7–48.

Kuran, T. (1997). *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press.

Laajaj, R., & Macours, K. (2019, October). Measuring Skills in Developing Countries. *Journal of Human Resources*, 1018-9805R1. doi: 10.3368/jhr.56.4.1018-9805R1

La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap operas and fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, *4*(4), 1–31. doi: 10.1257/app.4.4.1

Lane, T., Nosenzo, D., & Sonderegger, S. (2019). Law and Norms: Empirical Evidence. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3581720

Lantos, N. A., Kende, A., Becker, J. C., & McGarty, C. (2020). Pity for economically disadvantaged groups motivates donation and ally collective action intentions. *European Journal of Social Psychology*, *50*(7), 1478–1499. doi: 10.1002/ejsp.2705

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012, January). Sorting in Experiments with Application to Social Preferences. *American Economic Journal: Applied Economics*, *4*(1), 136–163. doi: 10.1257/app.4.1.136

Lowe, M. (2020). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. , 115.

Lyon, N. (2023, April). Value Similarity and Norm Change: Null Effects and Backlash to Messaging on Same-Sex Rights in Uganda. *Comparative Political Studies*, *56*(5), 694–725. doi: 10.1177/00104140221115173

Mal, S. (2015). Let Us to Live: Social Exclusion of Hijra Community. *Asian Journal of Research in Social Sciences and Humanities*, *5*(4), 108. doi: 10.5958/2249-7315.2015.00084.2

Mallett, R. K., Huntsinger, J. R., Sinclair, S., & Swim, J. K. (2008, October). Seeing Through Their Eyes: When Majority Group Members Take Collective Action on Behalf of an Outgroup. *Group Processes & Intergroup Relations*, *11*(4), 451–470. doi: 10.1177/1368430208095400

Masih, P., Singh, G., & Mishra, R. (2012). *Ummeed live 2012: Third gender leadership development project*. Raipur, Chhattisgarh.

McAdams, R. H. (2000). Focal Point Theory of Expressive Law. *VirginiaLaw Review*, *86*, 83.

McAdams, R. H. (2001). An Attitudinal Theory of Expressive Law. *SSRN Electronic Journal*. doi: 10.2139/ssrn.253331

McAdams, R. H., & Rasmusen, E. B. (2004). Norms in Law and Economics. *SSRN Electronic Journal*. doi: 10.2139/ssrn.580843

Morris, S. (2001, April). Political Correctness. *Journal of Political Economy*, *109*(2), 231–265. doi: 10.1086/319554

Muralidharan, K., Romero, M., & Wüthrich, K. (2023, March). Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments. *The Review of Economics and Statistics*, 1–44. doi: 10.1162/rest_a_01317

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, *83*, 602–627. doi: 10.1037/0033-2909.83.4.602

Nuttbrock, L. (2018). *Transgender Sex Work and Society*. Columbia University Press.

O'Driscoll, M. P., & Feather, N. T. (1985). Positive Prejudice in Ethnic Attitudes: Australian Data. *International Journal of Psychology*, *20*(1), 95–107. doi: 10.1002/j.1464-066X.1985.tb00016.x

Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019, April). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences*, *116*(18), 8846–8851. doi: 10.1073/pnas.1806000116

Paluck, E. L., Green, S. A., & Green, D. P. (2019, November). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*(02), 129–158. doi: 10.1017/bpp.2018.25

Park, A., Bryson, C., & Curtis, J. (2014). *British Social Attitudes 31.* NatCen London.

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology*, *49*(1), 65–85. doi: 10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. doi: 10.1037/0022-3514.90.5.751

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, *62*(4), 659–661.

Plant, E. A., & Devine, P. (2009, March). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, *96*(3), 640–652. doi: 10.1037/a0012960

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145. doi: 10.1037/a0033323

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*. doi: 10.1257/aer.20180044

Reddy, G. (2005). *With Respect to Sex: Negotiating Hijra Identity in South India*. Chicago, IL: University of Chicago Press.

Rodríguez, G., & Elo, I. (2003, March). Intra-class Correlation in Random-effects Models for Binary Data. *The Stata Journal*, *3*(1), 32–46. doi: 10.1177/1536867X0300300102

Sangama. (2015). *Transgender Survey Kerala 2014-15* (Tech. Rep.).

Sansone, D. (2018, April). *Pink Work: Same-Sex Marriage, Employment and Discrimination* (SSRN Scholarly Paper No. 3164515). Rochester, NY. doi: 10.2139/ssrn.3164515

Schwardmann, P., Tripodi, E., & van der Weele, J. J. (2022, April). Self-Persuasion: Evidence from Field Experiments at International Debating Competitions. *American Economic Review*, *112*(4), 1118–1146. doi: 10.1257/aer.20200372

Searle, S. R., & Gruber, M. H. (2016). *Linear models*. John Wiley & Sons.

Shaikh, S., Mburu, G., Arumugam, V., Mattipalli, N., Aher, A., Mehta, S., & Robertson, J. (2016). Empowering communities and strengthening systems to improve transgender health: Outcomes from the Pehchan programme in India. *Journal of the International AIDS Society*, *19*(3 Suppl 2), 20809. doi: 10.7448/IAS.19.3.20809

Sharma, D. C. (2014, June). Changing landscape for sexual minorities in India. *The Lancet*, *383*(9936), 2199–2200. doi: 10.1016/S0140-6736(14)61070-9

Shivakumar, S. T., & Yadiyurshetty, M. M. (2014). Markers of well-being among the hijras: The male to female transsexuals. In S. Cooper & K. Ratele (Eds.), *Psychology serving humanity: Proceedings of the 30th international congress of psychology* (Vol. 1, pp. 218–232). New York: Psychology Press.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, *94*(4), 718–737. doi: 10.1037/0022-3514.94.4.718

Stasser, G., & Davis, J. H. (1981, November). Group decision making and social influence: A social interaction sequence model. *Psychological Review*, *88*(6), 523–551. doi: 10.1037/0033-295X.88.6.523

Subramanian, T., Gupte, M., Dorairaj, V., Periannan, V., & Mathai, A. (2009, April). Psychosocial impact and quality of life of people living with HIV/AIDS in South India. *AIDS Care*, *21*(4), 473–481. doi: 10.1080/09540120802283469

Sunstein, C. R. (1996, May). On the Expressive Function of Law. *University of Pennsylvania Law Review*, *144*(5), 2021. doi: 10.2307/3312647

Sunstein, C. R. (2019). *How Change Happens*. MIT Press.

Tankard, M. E., & Paluck, E. L. (2017, September). The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes. *Psychological Science*, *28*(9), 1334–1344. doi: 10.1177/0956797617709594

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009, November). Transforming "apathy into movement": The role of prosocial emotions in motivating action for social change. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *13*(4), 310–333. doi: 10.1177/1088868309343290

Tilcsik, A. (2011). Pride and prejudice: Employment discrimination against openly gay men in the United States. *American Journal of Sociology*, *117*, 586–626. doi: 10.1086/661653

Tsai, J., Lee, C. Y. S., Byrne, T., Pietrzak, R. H., & Southwick, S. M. (2017). Changes in Public Attitudes and Perceptions about Homelessness Between 1990 and 2016. *American Journal of Community Psychology*, *60*(3-4), 599–606. doi: 10.1002/ajcp.12198

U.S. State Dept. (2021). *2021 Country Reports on Human Rights Practices: India* (Tech. Rep.).

Wheaton, B. (2020). Laws, Beliefs, and Backlash. , 102.

Young, A. (2019, May). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. *The Quarterly Journal of Economics*, *134*(2), 557–598. doi: 10.1093/qje/qjy029

A Additional tables and figures

Figure A1: *Experimental design (detailed)*

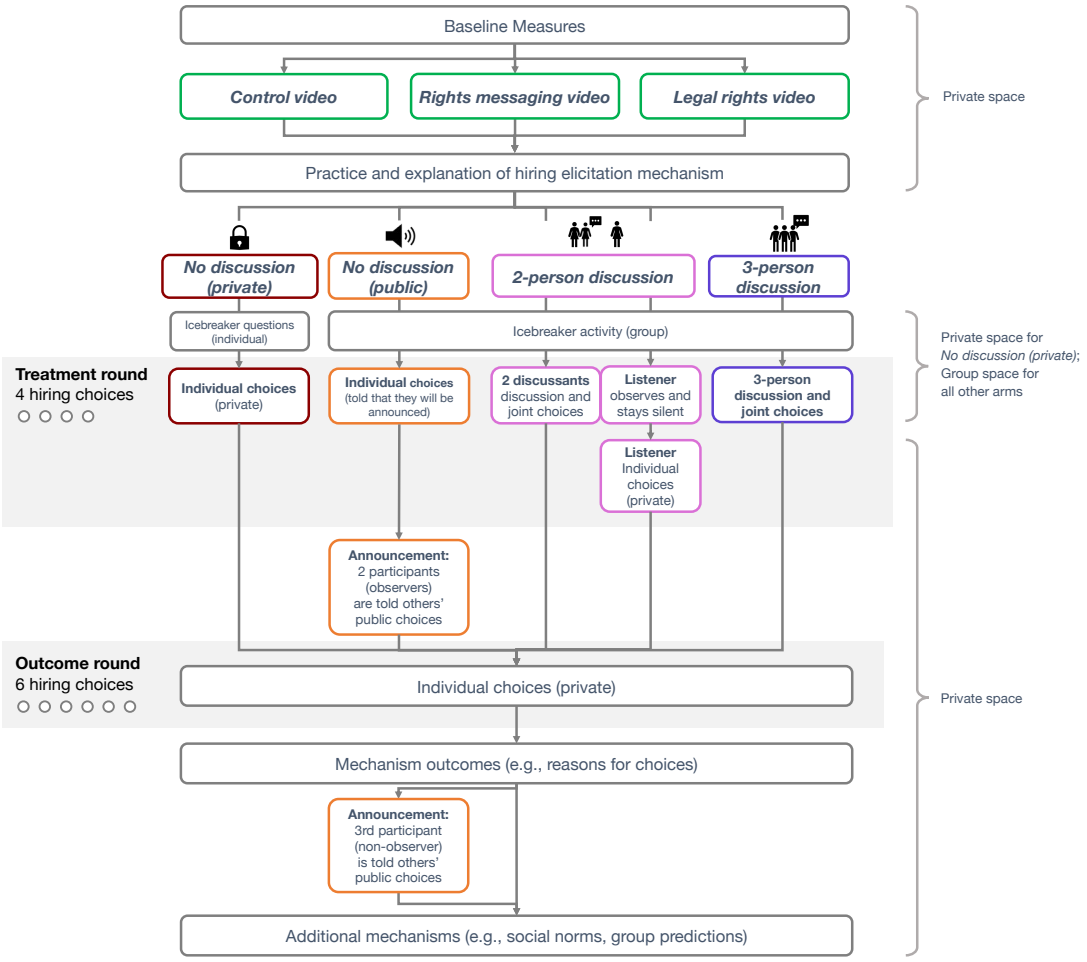
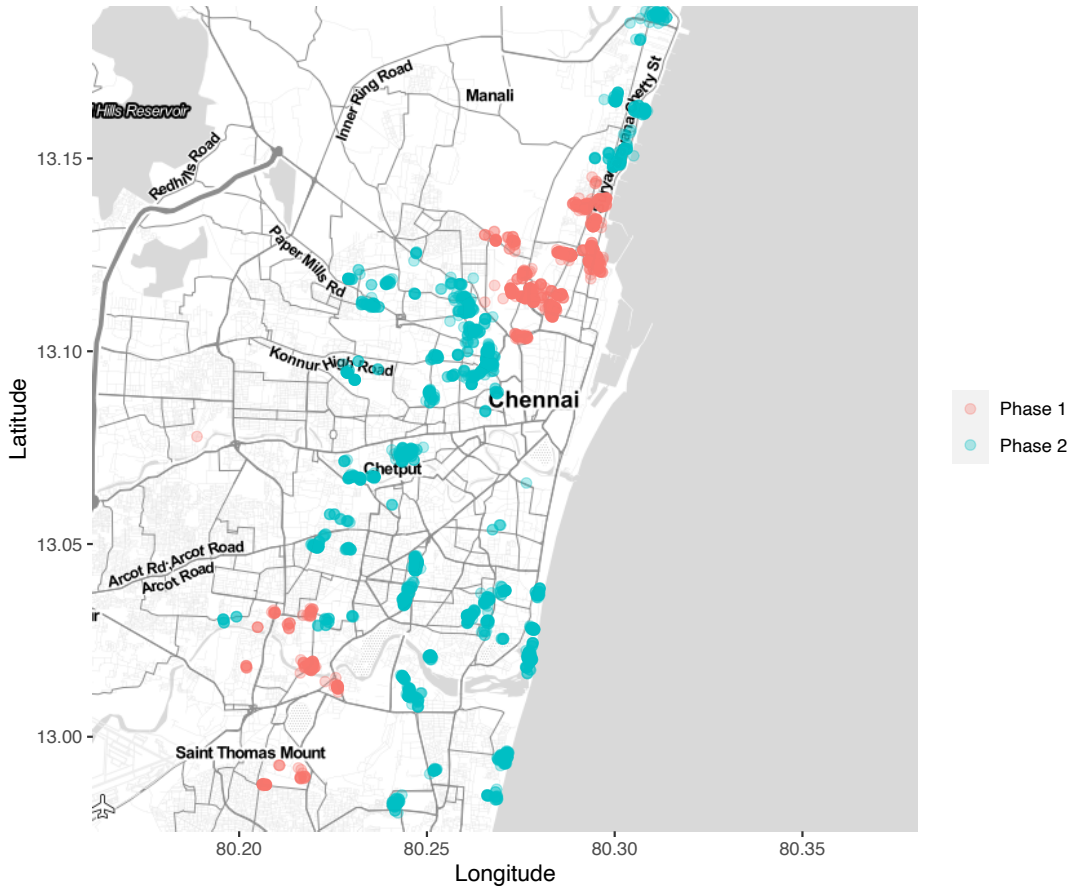


Figure A2: *Survey locations*



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: *3-person discussion results are robust to sampling weights that equalise weight of phase 1 and 2 across treatment conditions*

	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]	(0.022) [<0.001]	
Worker is trans	−0.193***		
	(0.013) [<0.001]		
3-person discussion	−0.004	0.002	0.167***
	(0.011) [0.729]	(0.010) [0.882]	(0.020) [<0.001]
Num. observations	13 494	13 494	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Mean: no discussion (private), worker is non-trans	0.61	0.61	
Mean: no discussion (private), worker is trans	0.42	0.42	0.42

Notes: All observations in the *3-person discussion* condition in phase 2 are given a relative weight of 2.30. This equalises the ratio of phase 1 and phase 2 observations across each treatment condition. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. 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. Column (3) only includes choices that involved a transgender worker. The specification used is seen in equation 1. 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 J.9). 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.

Table A5: *Balance for 3-person discussion (Phases 1 + 2)*

	Means		p-values	
	(1)	(2)		
Variable	No discussion (private)	3-person discussion	p-value (1)=(2)	
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 household food expenditure per capita (Rs.)	2310.04	2309.80	1.00	
F-test: statistic			**0.96**	
F-test: p-value			**0.48**	

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 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)
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 household food expenditure per capita (Rs.)	2304.84	2249.57	2345.54	2264.07	0.45	0.60	0.65
F-test: statistic					**0.83**	**0.50**	**0.38**
F-test: p-value					**0.62**	**0.91**	**0.97**

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: *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)
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 household food expenditure per capita (Rs.)	2373.21	2232.74	2278.81	0.05 *	0.18
F-test: statistic				**0.49**	**0.98**
F-test: p-value				**0.92**	**0.46**

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 A8: *Discussion effect is not correlated with whether audio recording was refused*

	Chose trans in outcome round (=1) (pairs with trans only)
	(1)
3-person discussion	0.168***
	(0.022) [<0.001]
3-person discussion × Audio recording refused (=1)	−0.009
	(0.045) [0.844]
Num. observations	4498
Num. participants	2249
Num. groups	751
Controls	X
Mean: Audio recording refused | 3-person discussion	0.14

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 the *3-person discussion* arm and the *No discussion (private)* arm, in both phases 1 and 2. *Audio recording refused* is coded as 0 for individuals in the *No discussion (private)* arm. Only choices that involved a transgender worker are included. The outcome is whether the transgender worker was selected. Controls include stratum fixed effects; dummies for the rights videos; whether the alternative worker was shown on the right; phase fixed effects; the relative # items offered by the alternative worker; the relative reliability score of the worker; a dummy for whether the reliability score was shown; and the controls selected by double LASSO (see Section J.9).

Table A9: *Logit model: discussion effect estimates are similar*

	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	0.159***	0.148***	
	(0.018) [<0.001]	(0.019) [<0.001]	
Worker is trans	−0.190***		
	(0.013) [<0.001]		
3-person discussion	−0.004	0.003	0.166***
	(0.011) [0.729]	(0.010) [0.754]	(0.020) [<0.001]
Num. observations	13 494	13 494	4498
Num. participants	2249	2249	2249
Num. participants	751	751	751
Controls		X	X
Controls interacted with worker is trans		X	

Notes: Coefficients are the average marginal treatment effects from a logit model. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets (not using randomization inference). 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*) 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 1. 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 J.9). 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 A10: *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***
	(0.021)	(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 dropping cases where the respondent said that others could hear their private outcome-round responses. Column 2 is the phase 2 sample, but dropping the cases when the listener spoke during the 2-person discussion. Column 3 is the phase 2 sample, but dropping the 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 J.9).

Table A11: *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
Mean: no discussion (private), worker is non-trans	0.62	0.62	
Mean: no discussion (private), worker is trans	0.37	0.37	0.37
Controls		X	X
Controls interacted with worker is trans		X	

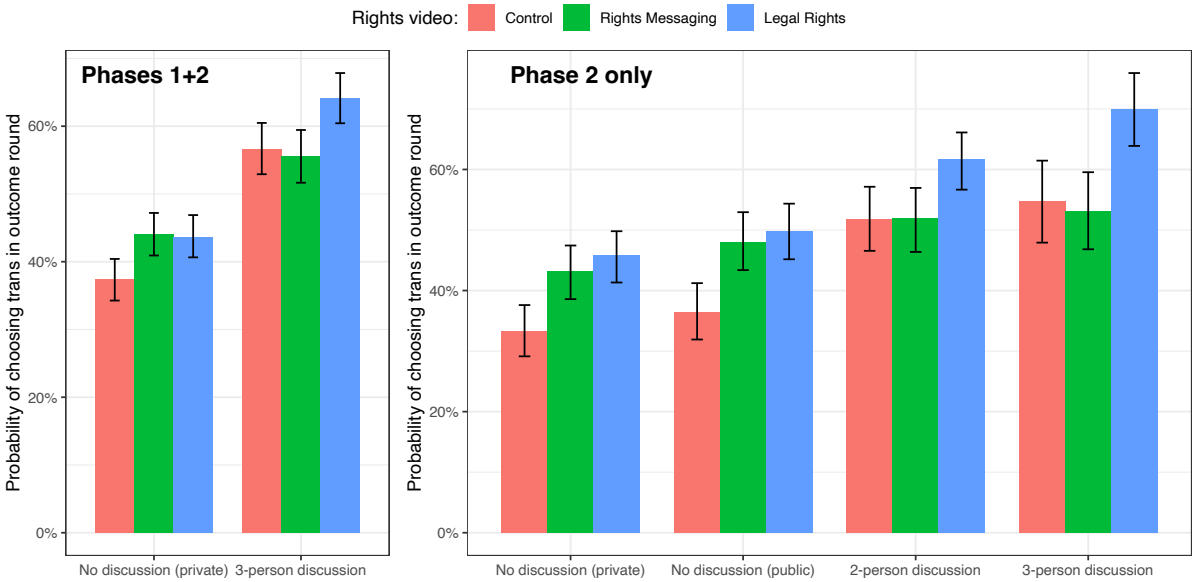
Notes: Sample includes *only* participants who saw the control video, and excludes participants who saw the rights messaging or legal rights videos. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in 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*) 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 1. 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 J.9). 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 A12: *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)	(3)	(4)
Rights messaging video	0.071**	0.070**	0.109***	0.110***
	(0.028) [0.012]	(0.028) [0.012]	(0.037) [0.003]	(0.037) [0.003]
Legal rights video	0.061**	0.060**	0.121***	0.119***
	(0.027) [0.025]	(0.027) [0.027]	(0.036) [<0.001]	(0.037) [0.001]
3-person discussion	0.194***	0.193***		
	(0.035) [<0.001]	(0.035) [<0.001]		
Rights messaging video × 3-person discussion	−0.082	−0.081		
	(0.051) [0.105]	(0.050) [0.109]		
Legal rights video × 3-person discussion	0.002	0.003		
	(0.048) [0.959]	(0.048) [0.957]		
No discussion (public)			0.029	0.026
			(0.039) [0.461]	(0.039) [0.502]
Discussion (pooled)			0.193***	0.189***
			(0.041) [<0.001]	(0.041) [<0.001]
Rights messaging video × No discussion (public)			0.007	0.008
			(0.058) [0.909]	(0.058) [0.889]
Rights messaging video × Discussion (pooled)			−0.111*	−0.109*
			(0.058) [0.056]	(0.058) [0.060]
Legal rights video × No discussion (public)			0.013	0.015
			(0.057) [0.815]	(0.057) [0.796]
Legal rights video × Discussion (pooled)			0.010	0.012
			(0.056) [0.860]	(0.056) [0.835]
Num. observations	4498	4498	4070	4070
Num. participants	2249	2249	2035	2035
Num. groups	751	751	741	741
Controls	X	X	X	X
p-val: (Rights messaging video | 3-person discussion)	0.804	0.822		
p-val: (Rights messaging video | No discussion (public))			0.009	0.008
p-val: (Rights messaging video | Discussion (pooled))			1.000	0.959
p-val: (Legal rights video | 3-person discussion)	0.099	0.100		
p-val: (Legal rights video | No discussion (public))			0.003	0.003
p-val: (Legal rights video | Discussion (pooled))			0.002	0.002

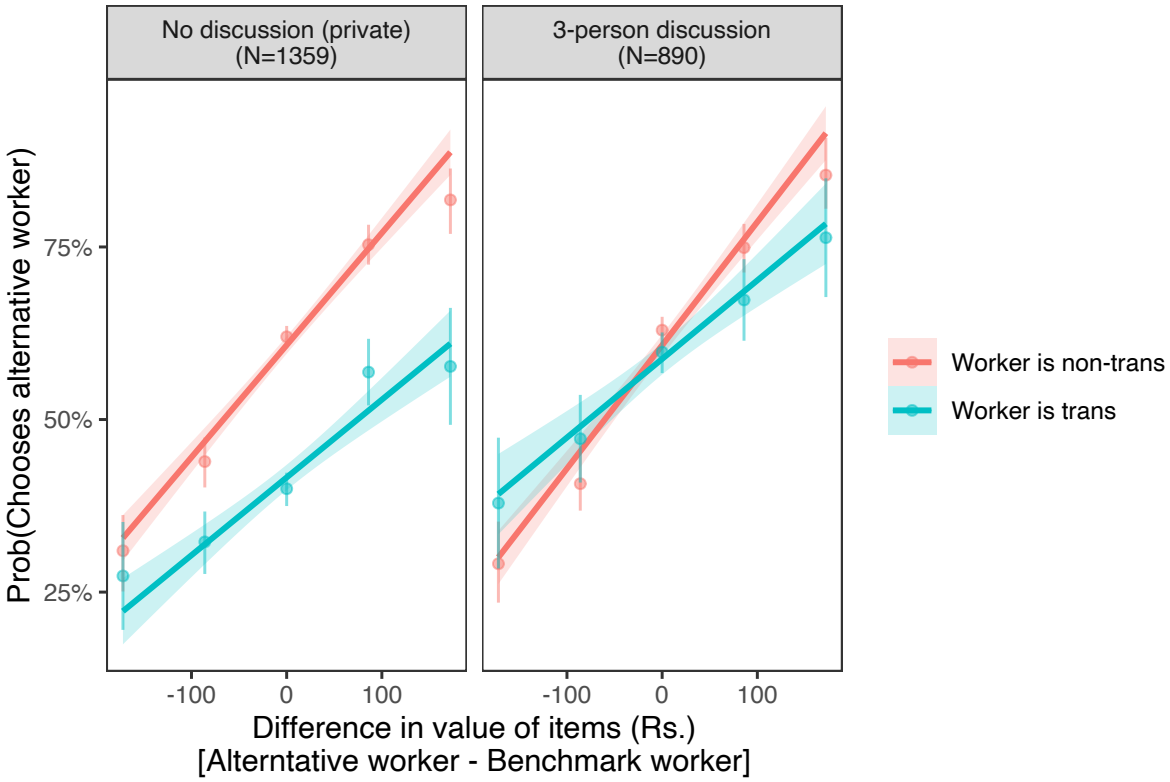
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Outcome is whether a participant chose the transgender worker in the private outcome 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. Columns (3) and (4) include participants in all discussion arms except *listeners* in the *2-person discussion* arm. *Discussion (pooled)* includes speakers in the *2-person discussion* arm and all participants in the *3-person discussion* arm. *p* (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 J.9).

Figure A13: *Interactions between rights videos and discussions*



Notes: 95% confidence intervals are based on standard errors clustered at the group-of-3 level. 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). Left panel includes only participants in the *No discussion (private)* and *3-person discussion* arms, in both phases 1 and 2. Right panel in includes all discussion-arm treatments in phase 2 of data collection.

Figure A14: *Inferring WTP to avoid transgender workers*



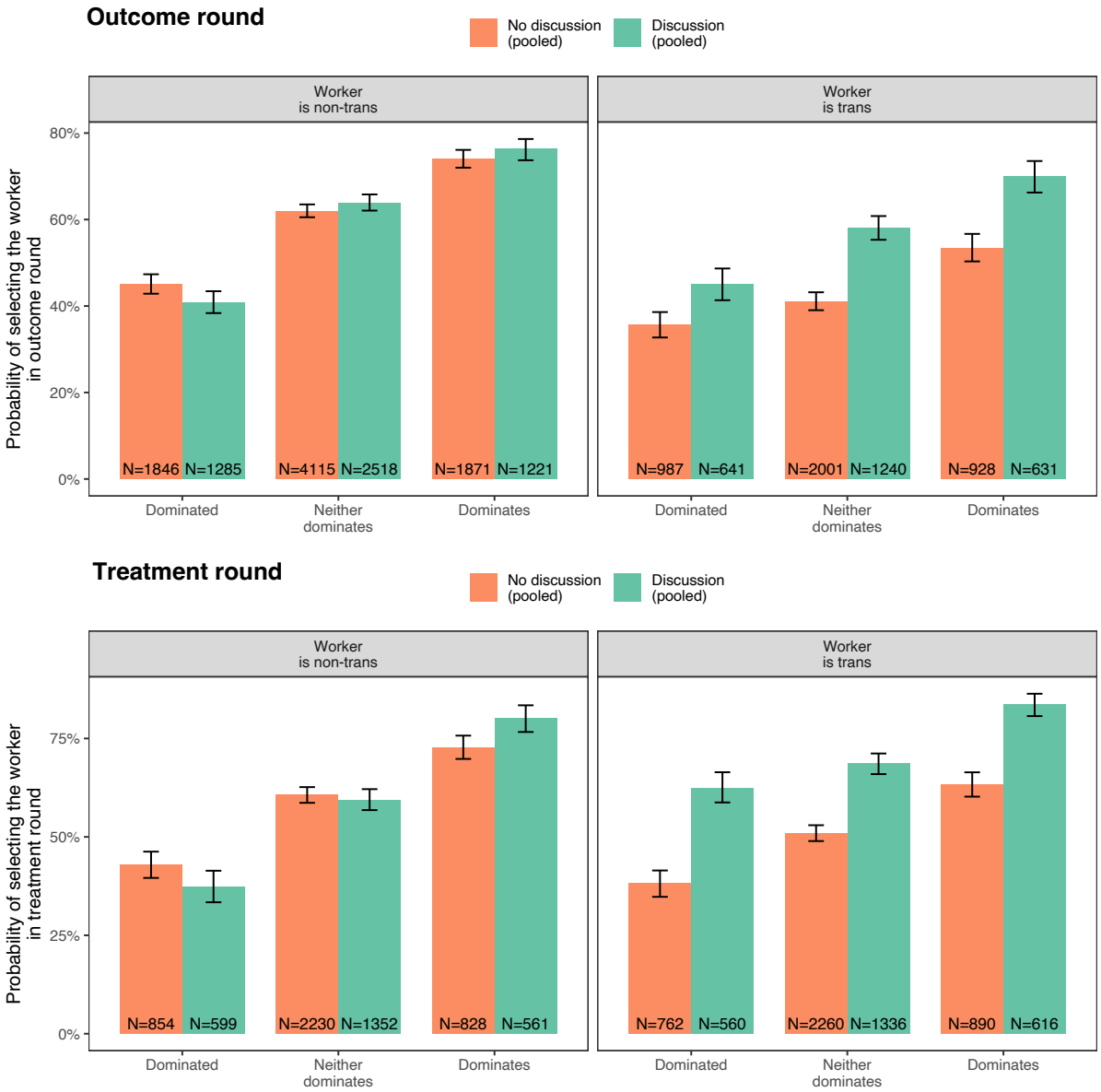
Notes: Points represent the probability of choosing the alternative worker at the given difference in value of items in Rs. Solid lines represent a linear fit. I take the reduction in probability that an option is chosen when a worker is transgender in each treatment group, and divide it by the gradient 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 A15: *Dominated and dominating choices: negative discrimination decreases and positive discrimination increases*

	Dep var: Chose worker (=1)	
	Outcome round	Treatment round
	(1)	(2)
Worker is trans	−0.190***	−0.097***
	(0.040) [<0.001]	(0.017) [<0.001]
Discussion (pooled)	0.017	−0.008
	(0.012) [0.166]	(0.024) [0.741]
Worker dominates	0.044***	0.046*
	(0.015) [0.003]	(0.027) [0.084]
Worker is dominated	−0.095***	−0.097***
	(0.016) [<0.001]	(0.027) [<0.001]
Worker is trans × Discussion (pooled)	0.150***	0.183***
	(0.024) [<0.001]	(0.034) [<0.001]
Worker is trans × Worker dominates	0.003	0.001
	(0.023) [0.897]	(0.028) [0.982]
Worker is trans × Worker is dominated	0.109***	0.048
	(0.023) [<0.001]	(0.031) [0.123]
Discussion (pooled) × Worker dominates	0.002	0.077*
	(0.020) [0.905]	(0.040) [0.053]
Discussion (pooled) × Worker is dominated	−0.054**	−0.059
	(0.023) [0.017]	(0.044) [0.183]
Worker is trans × Discussion (pooled) × Worker dominates	−0.006	−0.054
	(0.036) [0.866]	(0.054) [0.312]
Worker is trans × Discussion (pooled) × Worker is dominated	−0.020	0.111*
	(0.036) [0.589]	(0.060) [0.065]
Num. observations	19 284	12 848
Num. participants	3214	3212
Num. groups	1134	1133
Controls	X	X

Notes: Includes all participants from both phase 1 and 2, apart from listeners. *Discussion (pooled)* = participants in *3-person discussion* arm or speakers in the *2-person discussion* arm. *No discussion (pooled)* = participants in *No discussion (public)* or *No discussion (private)* arm. 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 participant × choice. Outcome is whether the participant selects the alternative worker instead of the male benchmark worker. Outcome round choices are on the top row, treatment round choices are on the bottom row. 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). *Dominated* is when the alternative worker is weakly dominated by the other option. *Dominates* is when the alternative worker weakly dominates the other option. *Neither dominates* is when neither the alternative 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; the relative number of items offered; the relative reliability score; whether the reliability score was shown; and the controls selected by double LASSO (see Section J.9).

Figure A16: *Dominated and dominating choices: negative discrimination decreases and positive discrimination increases*



Notes: Includes all participants from both phase 1 and 2, apart from listeners. *Discussion (pooled)* = participants in *3-person discussion* arm or speakers in the *2-person discussion* arm. *No discussion (pooled)* = participants in *No discussion (public)* or *No discussion (private)* arm. Unit of observation is participant × choice. Outcome is whether the participant selects the alternative worker instead of the male benchmark worker. Outcome round choices are on the top row, treatment round choices are on the bottom row.
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). *Dominated* is when the alternative worker is weakly dominated by the other option. *Dominates* is when the alternative worker weakly dominates the other option. *Neither dominates* is when neither the alternative worker nor the other option dominates.

Table A17: *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 1. 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 J.9). Controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

Table A18: *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

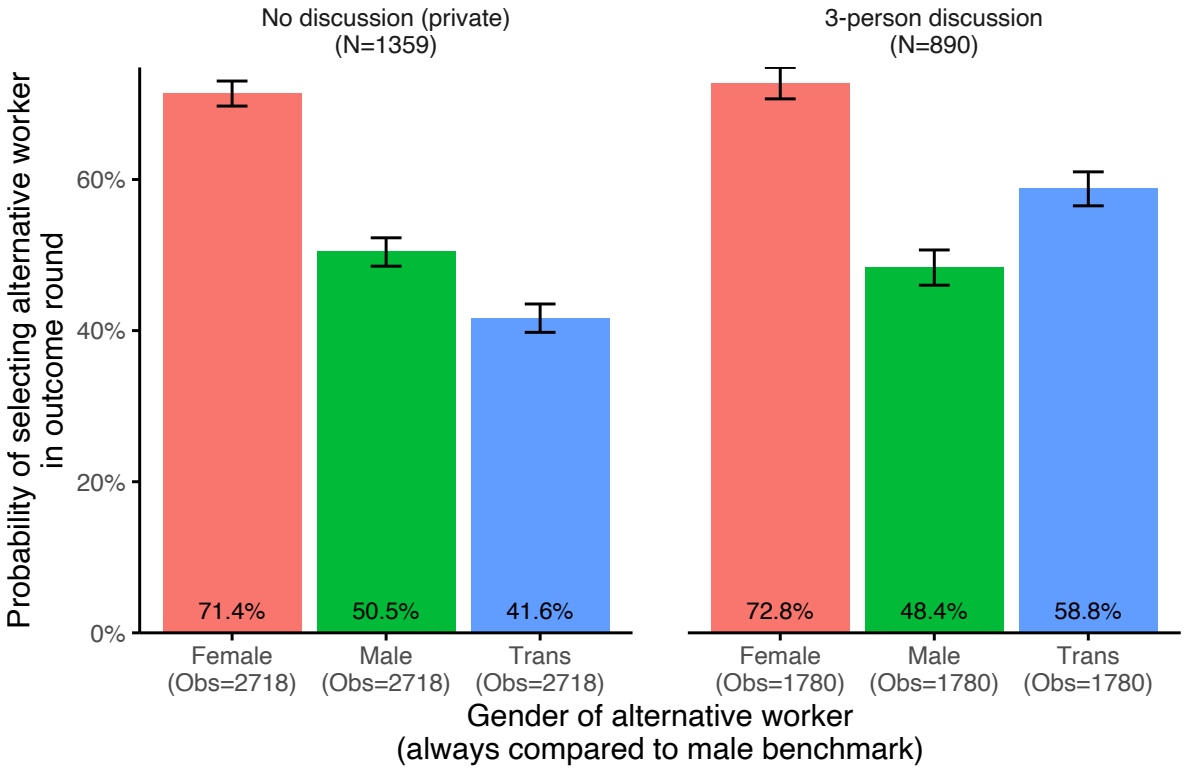
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The outcome is whether the *alternative worker* (rather than the male *benchmark worker*) 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 J.9). 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 A19: *Heterogeneity by demographic characteristics*

	Chose trans in outcome round (=1) (pairs with trans only)	
	Uninteracted term	Interacted term (x 3-person discussion)
	(1)	(2)
Female (=1)	0.068*	−0.004
	(0.038) [0.073]	(0.065) [0.951]
Speaks English (=1)	−0.016	0.037
	(0.040) [0.683]	(0.067) [0.577]
Reads English (=1)	−0.016	0.023
	(0.033) [0.614]	(0.052) [0.662]
Hindu (=1)	0.073**	−0.060
	(0.029) [0.013]	(0.049) [0.223]
Bachelor's degree (=1)	−0.018	−0.008
	(0.028) [0.526]	(0.049) [0.871]
Married (=1)	0.057*	−0.065
	(0.032) [0.071]	(0.053) [0.227]
Employed (=1)	0.055*	−0.096*
	(0.031) [0.082]	(0.050) [0.056]
Landlord (=1)	−0.017	0.088
	(0.037) [0.648]	(0.059) [0.137]
Has children (=1)	−0.002	−0.010
	(0.023) [0.928]	(0.038) [0.784]
Employer (=1)	−0.015	0.077*
	(0.025) [0.556]	(0.042) [0.066]
Above med. hh size (=1)	0.043*	−0.048
	(0.024) [0.072]	(0.038) [0.213]
Above med. hh food exp. p.c. (=1)	−0.002	0.012
	(0.022) [0.922]	(0.038) [0.760]
3-person discussion	0.279***	
	(0.086) [0.001]	
Num. observations	4452	4452
Num. participants	2249	2249
Num. groups	751	751
Controls	X	X

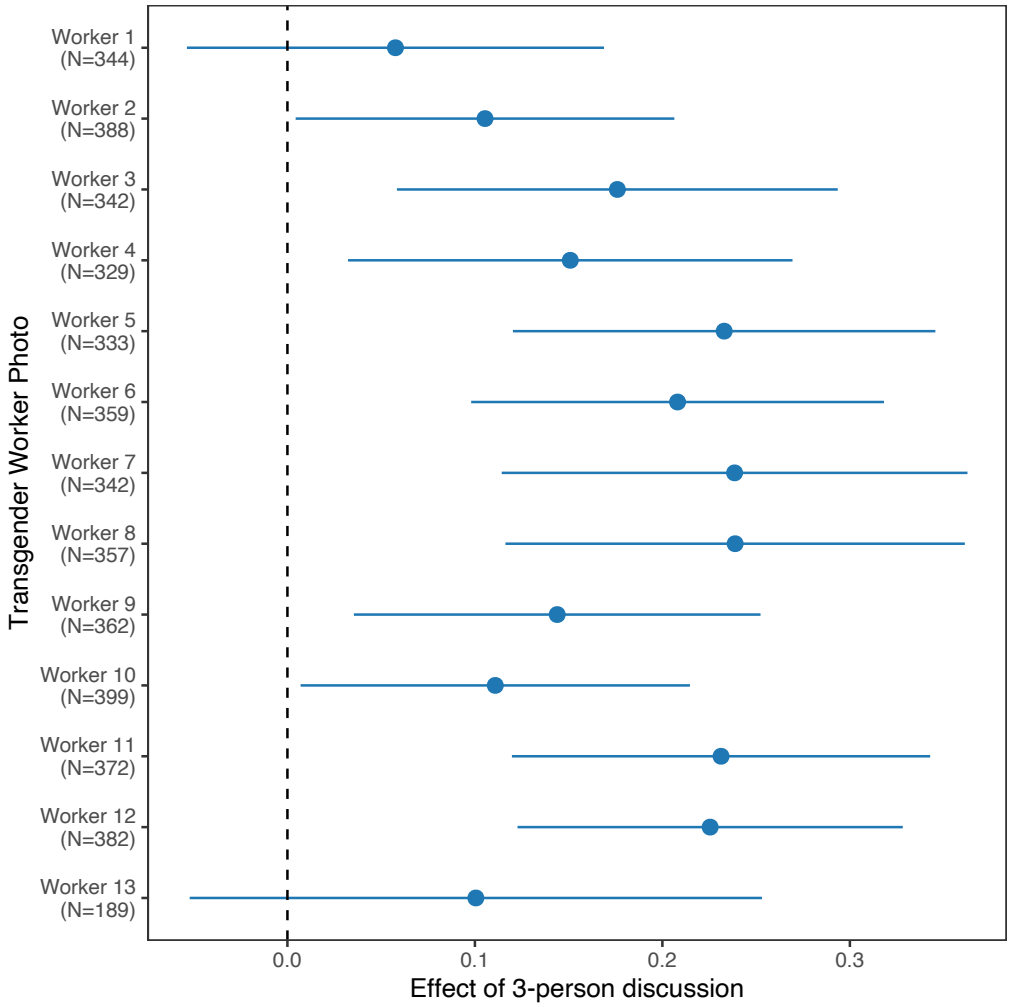
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, 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.

Figure A20: *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.

Figure A21: *Treatment effect of discussion is positive for all transgender worker photos*



Notes: Each plot shows the coefficient from the regression equivalent to Table 1, column 3, but restricting the sample to *only* the choices that included the worker photo on the vertical axis.

Table A22: *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
	(0.014) [0.525]	(0.020) [0.055]	(0.050) [<0.001]	(0.015) [0.802]	(0.018) [0.467]	(0.028) [0.218]
Legal rights video	0.098***	0.195***	0.890***	0.034**	0.034**	0.269***
	(0.011) [<0.001]	(0.018) [<0.001]	(0.054) [<0.001]	(0.014) [0.016]	(0.017) [0.044]	(0.026) [<0.001]
Num. participants	3397	3397	3397	3397	3397	3397
Num. groups	1134	1134	1134	1134	1134	1134
Mean: Control video	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 J.9). *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 (Anderson, 2008).

Table A23: *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
Mean: No discussion (private)	0.96	
Mean: Control video		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 A24: *No evidence of geographical spillovers on the medium run choices 2-9 weeks later*

	Dep var: Chose trans in follow-up round (=1)					
	Range: 200m		Range: 500m		Range: 1km	
	(1)	(2)	(3)	(4)	(5)	(6)
3-person discussion	0.061***	0.066***	0.054**	0.065***	0.051**	0.060**
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)	(0.023)
	[0.008]	[0.005]	[0.021]	[0.005]	[0.030]	[0.010]
No discussion (private) + Above med. number of discussion groups nearby	0.016		0.001		−0.005	
	(0.024)		(0.025)		(0.028)	
	[0.505]		[0.969]		[0.868]	
No discussion (private) + Above med. proportion of discussion groups nearby		0.025		0.023		0.013
		(0.024)		(0.025)		(0.025)
		[0.280]		[0.355]		[0.605]
Num. observations	4052	4032	4052	4050	4052	4050
Num. participants	2127	2127	2127	2127	2127	2127
Num. groups	745	745	745	745	745	745

Notes: The table measures for spillover effects in the follow-up round 2-9 weeks after the main data collection. *No discussion (private) + Above med. number of discussion groups nearby* is 1 if the participant was in the *No discussion (private)* arm *and* the number of other groups within a certain geographical range (given in the column header) who were in the *3-person* or *2-person* discussion arms is above the median. *No discussion (private) + Above med. proportion of discussion groups nearby* is 1 if the participant was in the *No discussion (private)* arm *and* the proportion of other groups within a certain geographical range (given in the column header) who were in the *3-person* or *2-person* discussion arms is above the median. All regressions control for the rights video and stratum fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets. Only participants in the *3-person discussion* and *No discussion (private)* arms are included. The omitted category is therefore *No discussion (private)* arm with below median number or proportion of discussion groups nearby. The dependent variable is whether the participant chose a transgender worker in the follow-up round 2-9 weeks after the main data collection.

Table A25: *Mediation analysis of discussion using mechanism outcomes*

	Dep var: Chose trans in private outcome round			
	(1)	(2)	(3)	(4)
3-person discussion	0.155***	0.068***	0.165***	0.163***
	(0.020)	(0.018)	(0.020)	(0.019)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Predicted probability of others choosing trans (community)	0.329***			
	(0.031)			
	[<0.001]			
Predicted probability of others choosing trans (group)		0.413***		
		(0.019)		
		[<0.001]		
Disagreed with discrimination (=1)			0.184***	
			(0.042)	
			[<0.001]	
Trans likely to complete delivery (=1)				0.229***
				(0.018)
				[<0.001]
Num. observations	4498	4476	4498	4498
Num. participants	2249	2249	2249	2249
Num. groups	751	751	751	751
Mean: No discussion (private)	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 $\times$ 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 Sections 3.9 and 5. 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 J.9).

Table A26: *Effect of rights video on predictions about others*

	Predicted % who pick trans (community)	Predicted % who pick trans (within group-of-3)
	(1)	(2)
Rights messaging video	0.023**	0.045**
	(0.011) [0.045]	(0.021) [0.033]
Legal rights video	0.027**	0.066***
	(0.011) [0.015]	(0.020) [0.001]
Num. observations	3397	6741
Num. participants	3397	3377
Num. groups	1134	1133
Controls	X	X
p(Rights messaging video=Legal rights video)	0.701	0.284
p(Legal rights video=3-person discussion)	0.326	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. 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 $\times$ 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 J.9).

Table A27: *Effect of trans rights videos on attitudes and beliefs*

	# 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	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]		
Rights messaging video		−0.006	
		(0.009) [0.464]	
Legal rights video		0.011	
		(0.008) [0.203]	
Photo is trans			−0.099***
			(0.030) [<0.001]
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	6794	6794	6794
Num. participants	3397	3397	3397
Num. groups	1134	1134	1134
Question FEs	X	X	X
Participant FEs	X		X
Discussion arm controls	X	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. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all participants in both phases.

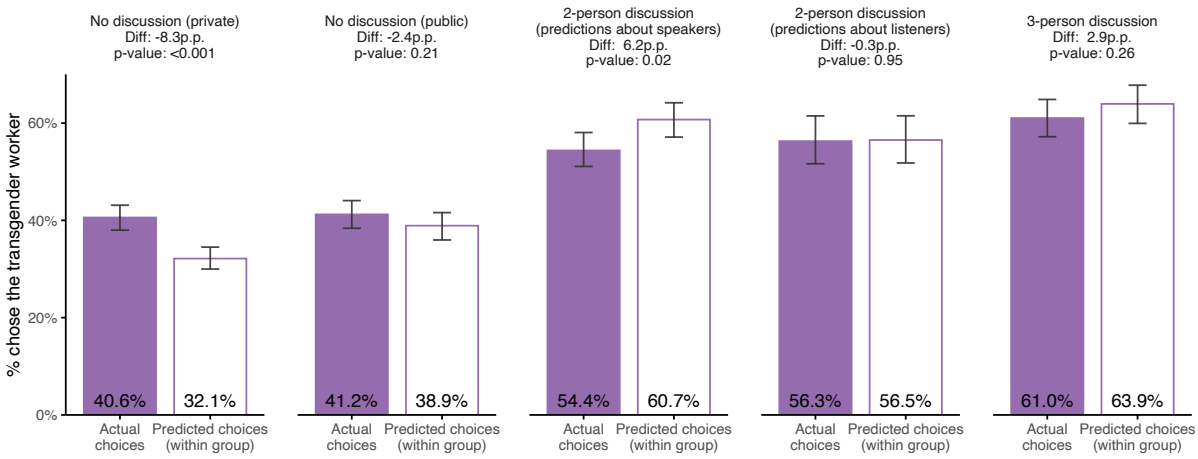
Column (1): dependent variable is the number of statements the participant agreed with on each of the lists of statements shown to them. Each participant sees both List A and List B, and the anti-trans statement ("In general, if I see a transgender person, I walk away") is randomly included in either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

Column (2): the enumerator describes two scenarios, one in which an employer rejects a transgender individual, and another in which a woman avoids a transgender individual on the street. The dependent variable is whether the participant says the person's actions are "wrong". *Question FEs* is a fixed effect for the second scenario.

Column (3): the dependent variable is whether the participant says a worker is likely or very likely to complete a delivery after being shown a photo of the worker. Participants make two choices each, one of which includes a transgender photo. The order is randomized. *Question FEs* controls for the order of the choice.

Controls include stratum fixed effects; dummies for the discussion-arm treatments; phase fixed-effects; and the controls selected by double LASSO (see Section J.9).

Figure A28: *Predictions about others in group (Phase 2)*



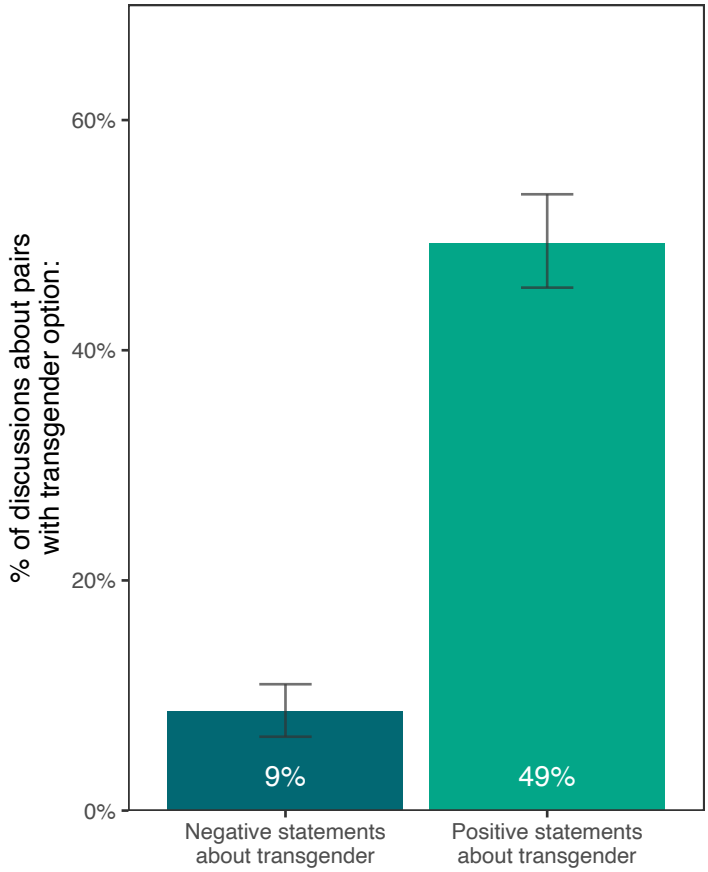
Notes: Sample includes all participants in phase 2. Unit of observation is participant × prediction. Only choices that include a transgender photo are included. Hollow bars represent the probability that a participant predicts that their group-member selects a transgender delivery worker. The prediction 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. Discussion speakers (pooling 2-person and 3-person discussions) are predicted to choose transgender workers 5.7 p.p. more than discussion listeners (p=0.04).

Table A29: *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]	(0.029) [<0.001]	
Worker is trans	−0.085***		
	(0.015) [<0.001]		
3-person discussion	0.007	0.004	0.199***
	(0.023) [0.741]	(0.021) [0.860]	(0.022) [<0.001]
Num. observations	8996	8996	4498
Num. participants	2249	2249	2249
Num. groups	751	751	751
Mean: no discussion (private), worker is non-trans	0.60	0.60	
Mean: no discussion (private), worker is trans	0.51	0.51	0.51
Controls		X	X
Controls interacted with worker is trans		X	

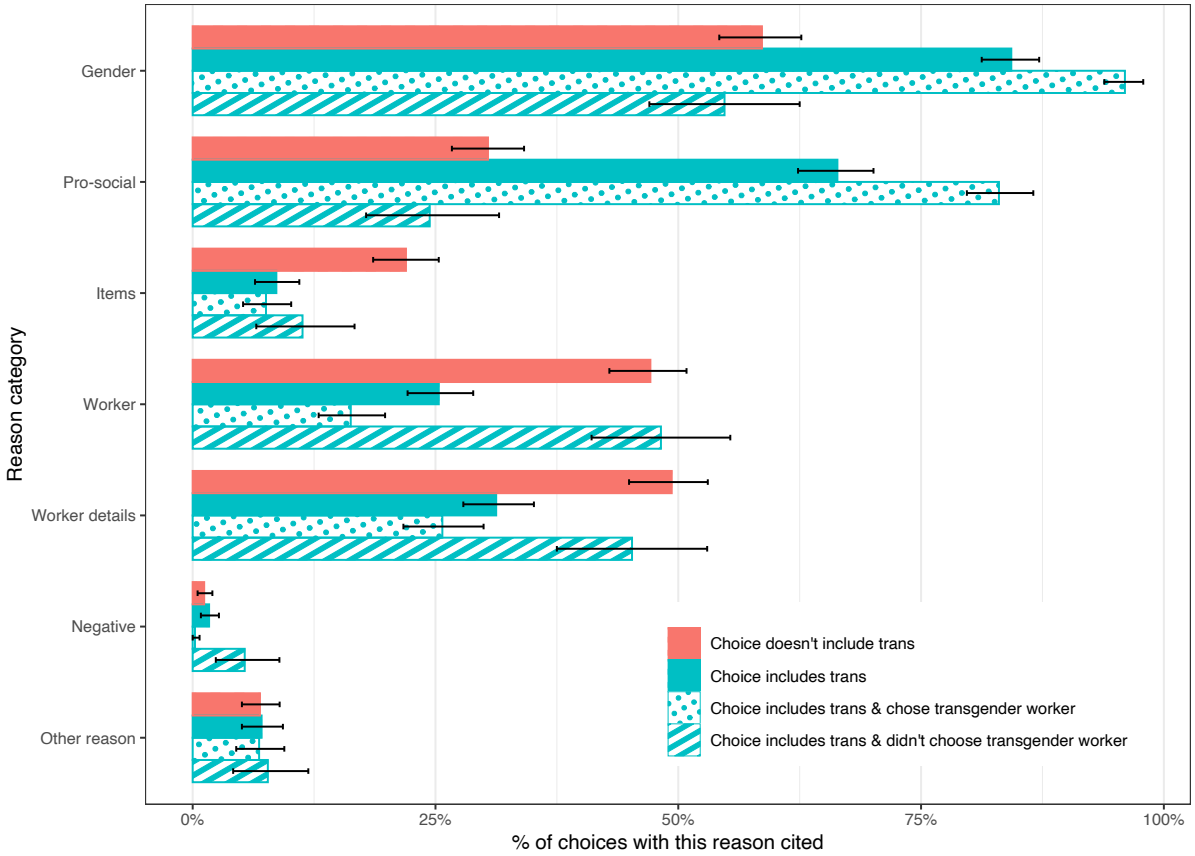
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Randomization inference p-values are in brackets. Sample includes the *3-person discussion* arm and the *No discussion (private)* arm, in both phase 1 and 2. The outcomes are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

Figure A30: *Participants in the discussion are more likely to say positive statements about transgender workers than negative statements*



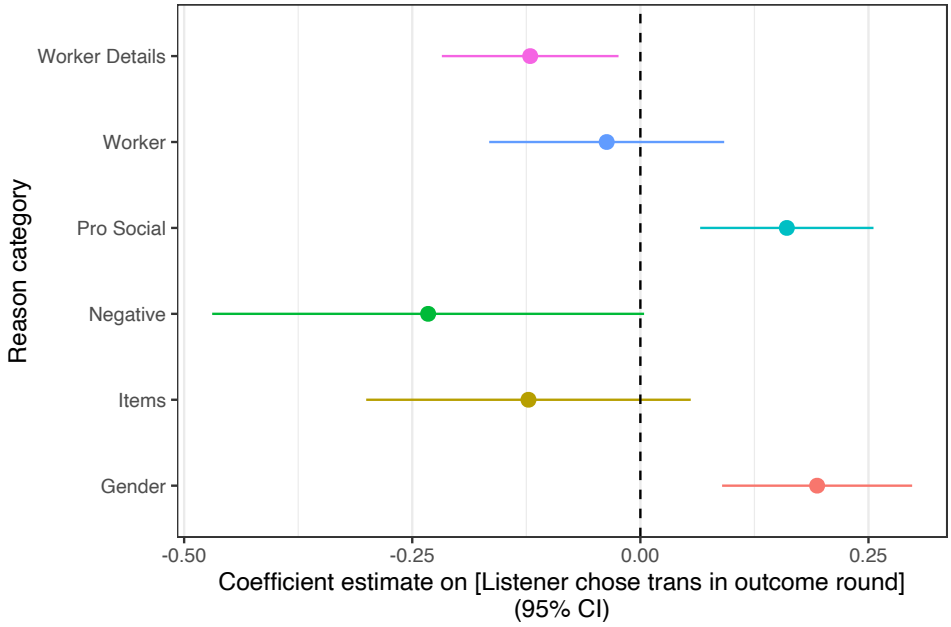
Notes: Unit of observation is the participant × choice level. Only choices that include a transgender worker are included. Enumerators coded for each choice whether each participant said a positive statement about the transgender worker, a negative statement about the worker, or both. Participants are 5.7x (= 49%/9%) more likely to say a positive statement rather than a negative statement about transgender workers in the discussion. Sample used is the *3-person discussion* arm only, in both phase 1 and phase 2.

Figure A31: *Reasons cited in the 3-person discussions*



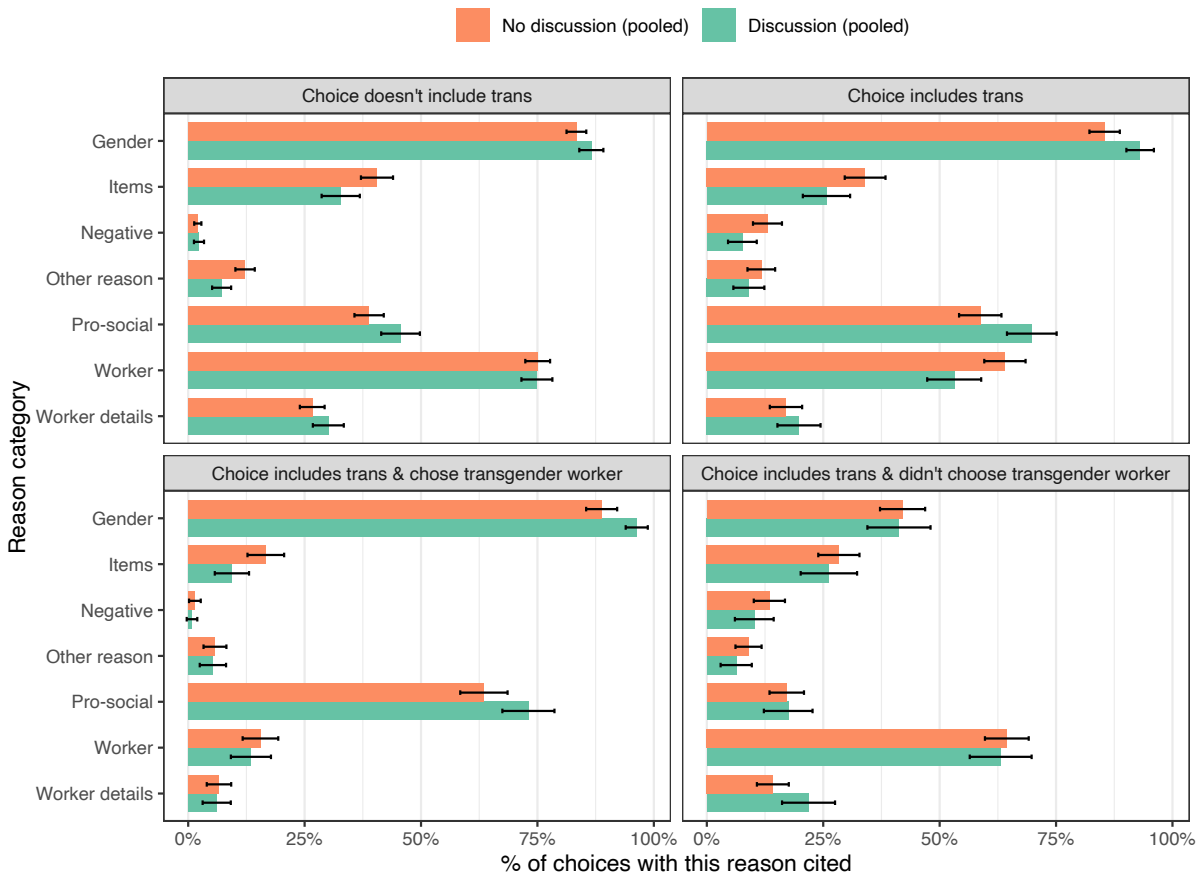
Notes: Unit of observation is a group × choice. Sample is the *3-person discussion* arm in both phase 1 and 2. Confidence intervals are based on a bootstrapped binomial distribution. One enumerator observed the discussion and marked the main reasons that the participants said they were selecting the 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 A32: *Correlation between reason cited and listener's choices*



Notes: This shows the correlation between the reasons a *Listener* heard in the discussion and their choices in the private outcome round. Reasons in each category are given in Figure A31. I regress the probability that a listener chose a transgender worker in the private outcome round on the number of choices for which they heard a specific reason category cited (restricting to only discussion choices that involved a transgender worker). I use a separate regression for each reason category. I control for the rights video seen by the participant, stratum fixed effects, whether the transgender worker was shown on the right, the relative number of items on offer, the relative reliability score, and whether it was shown. The sample includes only *Listeners* in the 2-person discussion arm. 95% confidence intervals are constructed using standard errors clustered at the individual level. Unit of observation is individual $\times$ choice.

Figure A33: *Retrospective reasons for choices in outcome round (aggregated)*



Notes: In phase 2, after participants finished the full set of hiring choices, they were asked *why* they chose 4 randomly selected options in the outcome round. Unit of observation is the participant × choice × reason. Outcome is whether the reason was given when asked why the participant made their choice for a given pair of options from the outcome round. *No discussion (pooled)* includes all participants in the *No discussion (private)* and *No discussion (public)* arms. *Discussion (pooled)* includes speakers in the *2-person discussion* and *3-person discussion* arms, and does not include listeners. Confidence intervals are calculated based on standard errors clustered at the group-of-3 level. Top left panel includes only choices that did not include a transgender worker. Top right panel includes only choices that included a transgender worker. Bottom left panel includes only choices where there was a transgender worker and the participant chose the transgender worker. Bottom right panel is only choices where there was a transgender worker and the transgender worker was not chosen. Reasons in each category are given in Figure A31.

Table A34: *Treatment round choices (Phase 2 sample)*

	Chose worker in treatment round (=1)		Chose trans in treatment round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans × Discussion participant	0.230***	0.224***	
	(0.036) [<0.001]	(0.033) [<0.001]	
Worker is trans × No discussion (public)	0.025	0.031	
	(0.033) [0.460]	(0.031) [0.319]	
Worker is trans	−0.097***		
	(0.021) [<0.001]		
Discussion participant	−0.021	−0.017	0.208***
	(0.026) [0.415]	(0.024) [0.483]	(0.024) [<0.001]
No discussion (public)	−0.008	−0.016	0.015
	(0.023) [0.715]	(0.021) [0.428]	(0.022) [0.478]
Num. observations	8132	8132	4066
Num. participants	2033	2033	2033
Num. groups	740	740	740
Mean: no discussion (private), worker is non-trans	0.59	0.59	
Mean: no discussion (private), worker is trans	0.49	0.49	0.49
Controls		X	X
Controls interacted with worker is trans		X	
p(No discussion (public)=Discussion participant)	0.000	0.000	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. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes all participants in phase 2 apart from 2-person discussion *listeners*. *Discussion participant* includes all participants the *3-person discussion* arm, and all speakers in the *2-person discussion* arm. *No discussion (public)* includes both non-observers and observers in that arm. The omitted category is *No discussion (private)*. Column (3) only includes choices that involved a transgender worker. In columns (1) and (2), the outcome is whether the *alternative worker* was selected (rather than the male *benchmark worker*) during the *treatment round* (i.e. during the discussion for those in a discussion arm). 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 1. 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 J.9). 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. Columns (2) and (3) also include controls for the relative number of items offered and the relative reliability score (which was always shown in the treatment round). randomization inference *p*-value at the base of the table tests for differences between the *No discussion (public)* and *Discussion participant* arms, i.e., for differences in the interacted terms in columns (1) and (2), and differences in the uninteracted terms in column (3).

Table A35: *Public treatment arm and discussions lead to convergence in behavior within a group*

Treatment	Sample	Round	ICC	95% CI	N groups	p-val: (a)=(b)
(a) No discussion (private)	Phase 2 only	Treatment	0.07	[0.03, 0.11]	253	0.06
(b) No discussion (public)			0.13	[0.08, 0.19]	199	
(a) No discussion (private)	Phase 2 only	Outcome	0.11	[0.07, 0.16]	253	0.12
(b) No discussion (public)			0.16	[0.11, 0.22]	200	
(a) No discussion (private)	Phases 1 + 2	Outcome	0.10	[0.07, 0.13]	454	0.00
(b) 3-person discussion			0.23	[0.19, 0.28]	297	

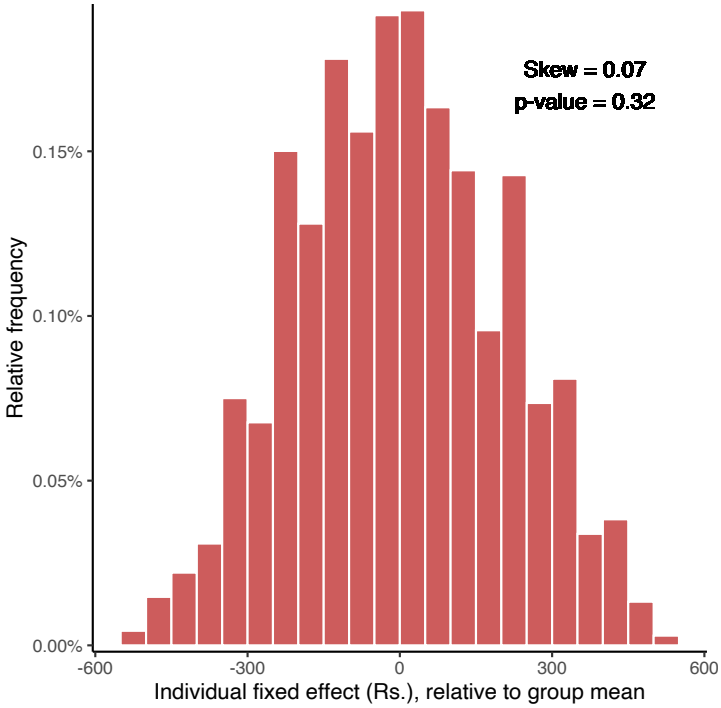
Notes: Unit of observation is the participant × choice. The variable of analysis is whether the participant selects a transgender worker, restricting to only choices where a transgender worker is shown. *ICC* denotes the intra-cluster correlation coefficient of this variable. The first two rows show results from choices in the treatment round. Rows 3-6 show results from choices in the private outcome round. In the treatment round, participants in a group are always shown the *same* options, regardless of their treatment option. In the outcome round, participants always see different options. 95% CI is calculated using the exact confidence limit equation from Searle & Gruber (2016). *p*-values are calculated using randomization inference that permutes the treatment status of each individual in the relevant treatment arms 1000 times.

Table A36: *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

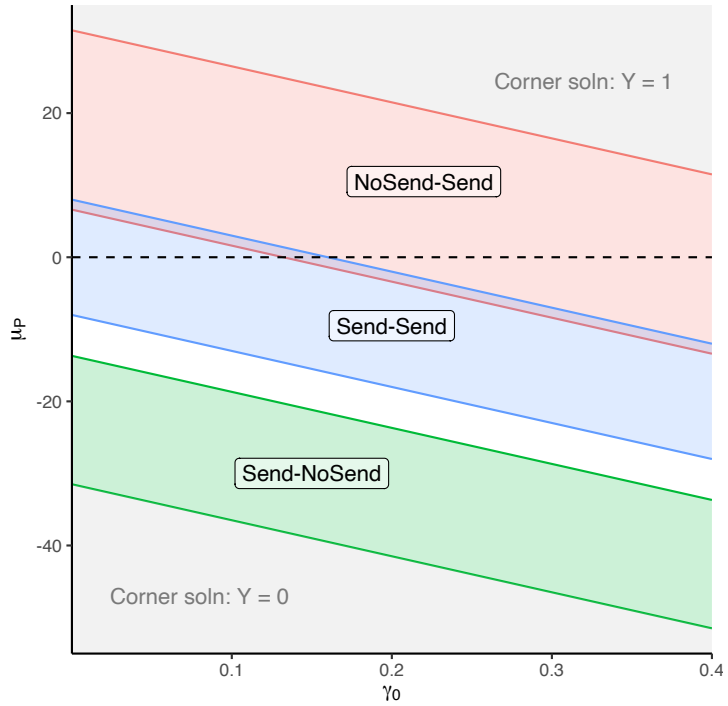
Notes: *P(spoke first)* is the probability that a participant spoke first in their group in the discussion of a choice, as marked by enumerator observations. The mean is 33%. *P(dominated)* is the probability that a participant dominated the discussion of a choice, as marked by enumerator observations. The mean is *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 × 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 A37: *The distribution of preferences for transgender workers within a group in the No discussion (private) arm is symmetric*



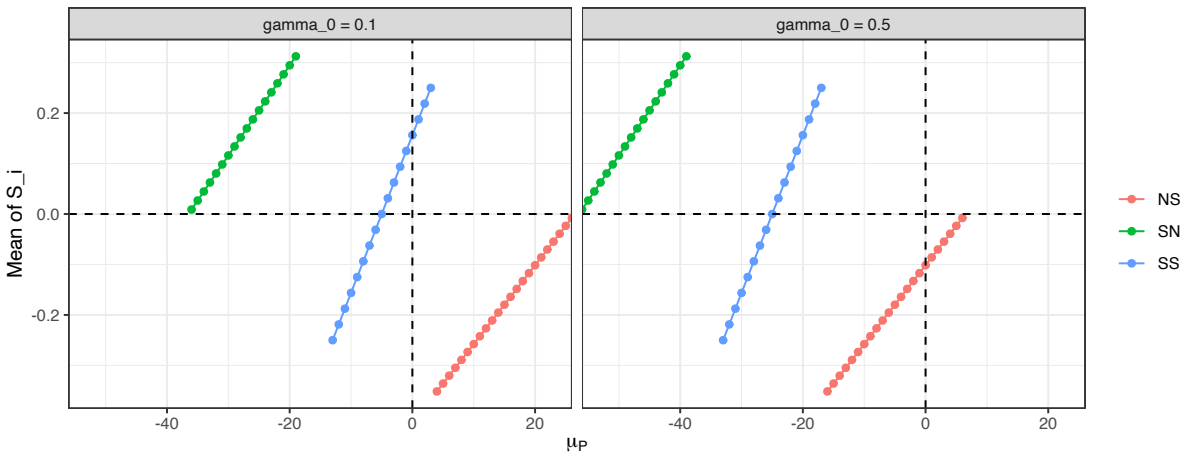
Notes: Sample is *No discussion (private)* only. I run a regression of whether the participant selects a transgender worker on a series of individual-specific fixed effects interacted with whether the choice included a transgender worker. The coefficient on these interaction terms gives an indication of the individual-specific preference for or against a transgender worker. I also control for the relative item value, reliability score, and whether the reliability score is shown in this regression. I divide the interacted fixed effects by the coefficient on relative item value to get the fixed effects in terms of monetary value (Rs.). The horizontal axis shows the value of the individual fixed effect, subtracted from the average fixed effect in a group of 3.

Figure A38: *Regions of parameter space for each equilibria in the model*



Notes: Parameter values: $\alpha = 10$, $\gamma_1 = 0.01$, $R = 50$, $c = 2.5$. Shows the range of values for (μ_P, γ_0) where each of the three equilibria in *Proposition 2* are feasible. The northeast region in gray is where only a corner solution in which everyone selects $Y_i = 1$ is feasible, and everyone chooses $S_i = 0$. The southwest region in gray is also corner solution where everyone selects $Y_i = 0$ and $S_i = 0$

Figure A39: *Net persuasion in the model*



Notes: Parameter values: $\alpha = 10$, $\gamma_1 = 0.01$, $R = 50$, $c = 2.5$. Horizontal axis shows the value of μ_P. Vertical axis shows the mean value of S_i, which depends on the type of equilibrium $Q \in \{SS, NS, SN\}$ and the proportion of people who choose Y_i in that equilibrium. The left panel shows results for $\gamma_0 = 0.1$, and the right panel shows results for $\gamma_0 = 0.5$.

Table A40: *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)
	(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	0.008	0.125***
	(0.023) [0.639]	(0.023) [0.738]	(0.030) [<0.001]
Worker is trans × Listener (2-person discussion)	0.113***	0.114***	
	(0.042) [0.007]	(0.041) [0.006]	
Listener (2-person discussion)	0.022	0.017	0.135***
	(0.027) [0.415]	(0.026) [0.528]	(0.033) [<0.001]
Num. observations	5012	5012	2506
Num. participants	1253	1253	1253
Num. groups	541	541	541
Mean: no discussion (private), worker is non-trans	0.63	0.63	
Mean: no discussion (private), worker is trans	0.34	0.34	0.34
Controls		X	X
Controls interacted with worker is trans		X	

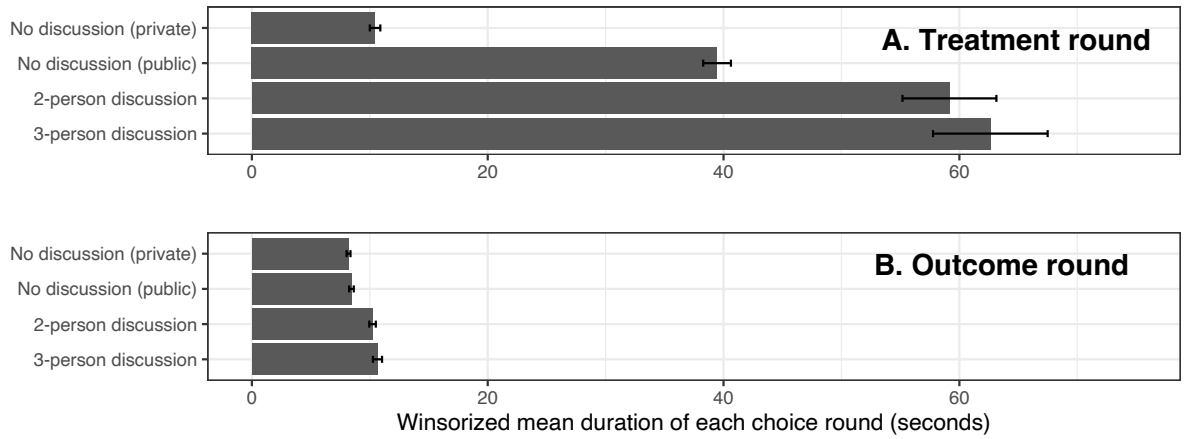
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. Standard p-values are in brackets. Unit of observation is the participant × choice level. Sample includes participants in phase 2 in the *3-person discussion* arm, the *No discussion (private)* arm, the listeners from the *2-person discussion* arm. 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*) 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 1. 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 J.9). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown.

Table A41: *Effect of transgender rights videos 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)
	(1)	(2)	(3)
Worker is trans	−0.281***		
	(0.020) [<0.001]		
Rights messaging video	0.013	0.014	0.057**
	(0.017) [0.469]	(0.017) [0.386]	(0.023) [0.012]
Legal rights video	0.020	0.019	0.091***
	(0.018) [0.255]	(0.017) [0.258]	(0.023) [<0.001]
Worker is trans × Rights messaging video	0.046	0.043	
	(0.028) [0.101]	(0.027) [0.121]	
Worker is trans × Legal rights video	0.080***	0.074***	
	(0.029) [0.006]	(0.027) [0.007]	
Num. observations	8872	8872	4436
Num. participants	2218	2218	2218
Num. groups	741	741	741
Controls		X	X
Controls interacted with worker is trans		X	
p(Rights messaging video=Legal rights video)	0.227	0.250	0.145

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 all discussion-arm treatments in phases 1 and 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*) 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. Controls include stratum fixed effects; dummies for the discussion-arm treatments; phase fixed-effects; and the controls selected by double LASSO (see Section J.9). In column (2), controls are interacted with *Worker is trans*, so the coefficient on *Worker is trans* is not shown. The *p*-value for the difference in treatment effect between the rights messaging and legal rights video is shown at the base of the table, which uses the terms interacted with *Worker is trans* in columns 1 and 2, and the uninteracted terms in column 3.

Figure A42: *Participants in discussion arms take longer to make choices*



Notes: Duration is the number of seconds between the start of a choice and the end of a choice. Duration is winsorized at the 99% level. Confidence intervals are based on standard errors clustered at the group-of-3 level. Duration of each choice was measured only in phase 2, so only phase 2 is included for analysis.

Table A43: *Correlation between duration of response and worker selection*

	Dep var: Chose worker (=1)			
	Treatment round		Outcome round	
	(1)	(2)	(3)	(4)
Worker is trans	−0.088***	−0.085**	−0.146***	−0.205***
	(0.024) [<0.001]	(0.036) [0.019]	(0.018) [<0.001]	(0.029) [<0.001]
Duration (mins)	−0.025	−0.106	−0.002**	−0.001
	(0.027) [0.357]	(0.117) [0.363]	(0.001) [0.020]	(0.001) [0.468]
Worker is trans × Duration (mins)	0.103***	−0.009	0.001	0.000
	(0.033) [0.002]	(0.145) [0.952]	(0.002) [0.431]	(0.003) [0.979]
No discussion (public)		0.091		−0.003
		(0.094) [0.331]		(0.020) [0.866]
2-person discussion		0.029		0.035
		(0.056) [0.610]		(0.022) [0.110]
3-person discussion		0.107		−0.002
		(0.101) [0.290]		(0.025) [0.938]
Worker is trans × No discussion (public)		−0.109		0.026
		(0.106) [0.304]		(0.044) [0.544]
Worker is trans × 2-person discussion		0.151**		0.117**
		(0.069) [0.028]		(0.048) [0.015]
Worker is trans × 3-person discussion		0.113		0.213***
		(0.120) [0.348]		(0.054) [<0.001]
Duration (mins) × No discussion (public)		−0.021		0.000
		(0.178) [0.908]		(0.002) [0.932]
Duration (mins) × 2-person discussion		0.070		−0.002
		(0.124) [0.575]		(0.002) [0.184]
Duration (mins) × 3-person discussion		0.010		0.001
		(0.146) [0.945]		(0.002) [0.761]
Worker is trans × Duration (mins) × No discussion (public)		0.156		0.003
		(0.208) [0.454]		(0.004) [0.535]
Worker is trans × Duration (mins) × 2-person discussion		0.003		0.001
		(0.154) [0.982]		(0.004) [0.889]
Worker is trans × Duration (mins) × 3-person discussion		0.097		−0.003
		(0.180) [0.588]		(0.005) [0.494]
Num. observations	6648	6648	13 308	13 308
Num. participants	2216	2216	2218	2218
Num. groups	740	740	741	741
Controls	X	X	X	X

Notes: Duration is the number of seconds between the start of a choice and the end of a choice. Duration is winsorized at the 99% level. Confidence intervals are based on standard errors clustered at the group-of-3 level. Duration of each choice was measured only in phase 2, so only phase 2 is included for analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. 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. Columns 1 and 2 show the effect on choices in the treatment round. Columns 3 and 4 show effects on choices in the outcome round. The dependent variable is whether the alternative worker was chosen. *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; relative number of items; relative reliability score; whether the relative reliability score was shown.

Table A44: *Discussion has no effect on other photo characteristics, conditional on photo gender*

	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. They also rated photos as to 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 is the *difference* 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 A45: *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) [0.534]	(0.013) [0.608]
Proportion of non-trans words remembered	0.189**		
	(0.080) [0.019]		
Num. observations	1179	2249	2249
Num. participants	1179	2249	2249
Num. groups	393	751	751
Mean: No discussion (private)	0.75	0.08	0.12
Mean: 3-person discussion	0.71	0.09	0.13
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 J.4, 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 class 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 J.9).

Table A46: *Treatment effect is not driven by people who correctly guess 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 Crowne & Marlowe (1960) index, described in Section J.3. *Column (4).* Participants were read two lists of words, described in Section J.4, 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 J.9).

Table A47: *Effect of rights videos on salience of transgender and perceived purpose of the experiment*

	Remembered word 'transgender' (=1) (Phase 1 only)	Correctly guess purpose (=1) (after main outcome)	Correctly guess purpose (=1) (end of experiment)
	(1)	(2)	(3)
Rights messaging video	0.016	0.050***	0.068***
	(0.030) [0.591]	(0.011) [<0.001]	(0.013) [<0.001]
Legal rights video	0.043	0.051***	0.070***
	(0.030) [0.143]	(0.011) [<0.001]	(0.013) [<0.001]
Proportion of non-trans words remembered	0.189**		
	(0.080) [0.019]		
Num. participants	1179	3397	3397
Num. groups	393	1134	1134
Mean: Control video	0.71	0.05	0.08
Controls	X	X	X

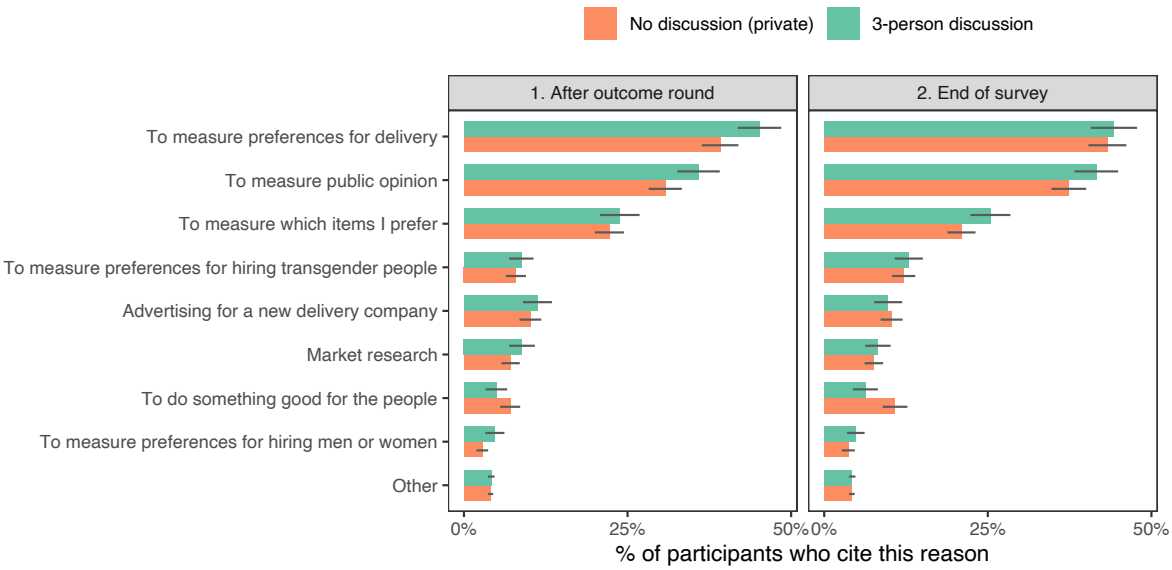
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. p-values are in brackets, and use randomization inference for the rights video coefficients. Unit of observation is the participant level. Sample includes participants from all discussion-treatment arms. 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 J.4, 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 class 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 discussion-arm treatments; and controls selected by double LASSO (see Section J.9).

Figure A48: *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 A49: *No significant differences in effect of the rights videos for participants who correctly guess the purpose, have high SDB score, or remember the word "transgender".*

	Chose trans in private outcome round (=1)			
	Phases 1 + 2		Phase 1 only	
	(1)	(2)	(3)	(4)
Rights messaging video	0.044**	0.049**	−0.026	−0.039
	(0.020) [0.028]	(0.020) [0.016]	(0.056) [0.638]	(0.059) [0.504]
Legal rights video	0.081***	0.079***	−0.012	−0.058
	(0.019) [<0.001]	(0.020) [<0.001]	(0.054) [0.829]	(0.057) [0.309]
Correctly guessed purpose (after main outcome)	0.049			
	(0.117) [0.678]			
Rights messaging video × Correctly guessed purpose (after main outcome)	−0.021			
	(0.065) [0.742]			
Legal rights video × Correctly guessed purpose (after main outcome)	−0.072			
	(0.063) [0.255]			
Correctly guessed purpose (end of experiment)		0.031		
		(0.045) [0.500]		
Rights messaging video × Correctly guessed purpose (end of experiment)		−0.047		
		(0.054) [0.386]		
Legal rights video × Correctly guessed purpose (end of experiment)		−0.040		
		(0.054) [0.462]		
Above median SDB score			−0.062	
			(0.042) [0.145]	
Rights messaging video × Above median SDB score			0.071	
			(0.061) [0.244]	
Legal rights video × Above median SDB score			0.030	
			(0.061) [0.627]	
Transgender word remembered				0.002
				(0.042) [0.962]
Above median proportion of non-trans words remembered				0.002
				(0.024) [0.927]
Rights messaging video × Transgender word remembered				0.091
				(0.068) [0.184]
Legal rights video × Transgender word remembered				0.091
				(0.062) [0.142]
Num. observations	6794	6794	2358	2358
Num. participants	3397	3397	1179	1179
Num. groups	1134	1134	393	393
Controls	X	X	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. *Column (3)*. SDB score is the social desirability score based on the Crowne & Marlowe (1960) index, described in Section J.3. *Column (4)*. Participants were read two lists of words, described in Section J.4, 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; relative # of items offered; relative reliability score; a dummy for whether the reliability score is shown; and controls selected by double LASSO (see Section J.9).

Table A50: *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	0.209***
	(0.032) [0.163]	(0.029) [0.325]	(0.048) [<0.001]
3 deliveries	0.030	0.032	−0.051
	(0.033) [0.374]	(0.028) [0.254]	(0.048) [0.293]
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	−0.056
	(0.044) [0.667]	(0.040) [0.904]	(0.074) [0.450]
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	

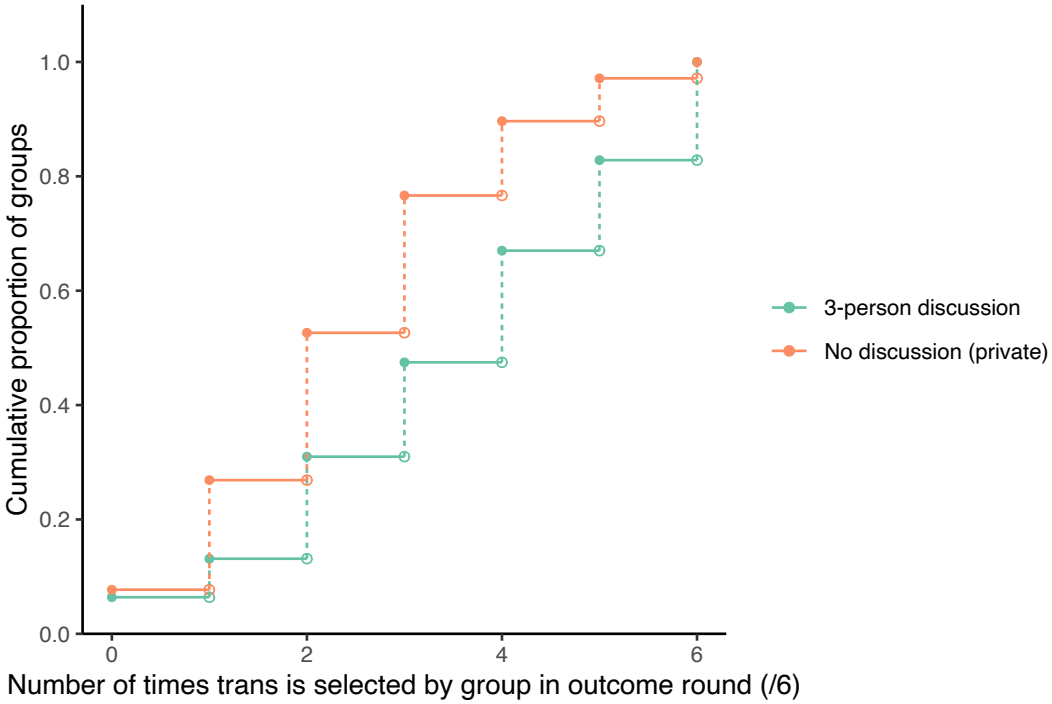
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the group-of-3 level and are in parentheses. randomization inference p-values are in brackets. Unit of observation is the participant × choice level. Sample includes only the subsample of 582 individuals in phase 1 who 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*) 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 1. 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 J.9). 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 A51: *Effect of rights videos may be attenuated with higher stakes*

	Chose worker in outcome round (=1)		Chose trans in outcome round (=1) (pairs with trans only)
	(1)	(2)	(3)
Worker is trans	−0.065		
	(0.059) [0.272]		
Rights messaging video	−0.010	−0.006	0.155**
	(0.044) [0.818]	(0.038) [0.881]	(0.063) [0.015]
Legal rights video	−0.008	−0.001	0.037
	(0.037) [0.828]	(0.032) [0.976]	(0.056) [0.516]
3 deliveries	−0.015	0.003	−0.058
	(0.037) [0.681]	(0.031) [0.916]	(0.062) [0.356]
Worker is trans × Rights messaging video	0.140*	0.171**	
	(0.080) [0.081]	(0.075) [0.024]	
Worker is trans × Legal rights video	0.045	0.043	
	(0.073) [0.533]	(0.068) [0.526]	
Worker is trans × 3 deliveries	−0.058	−0.057	
	(0.077) [0.458]	(0.071) [0.426]	
Rights messaging video × 3 deliveries	0.049	0.032	−0.062
	(0.057) [0.394]	(0.050) [0.526]	(0.094) [0.509]
Legal rights video × 3 deliveries	0.063	0.048	−0.008
	(0.051) [0.226]	(0.043) [0.258]	(0.084) [0.921]
Worker is trans × Rights messaging video × 3 deliveries	−0.092	−0.102	
	(0.113) [0.413]	(0.106) [0.335]	
Worker is trans × Legal rights video × 3 deliveries	−0.074	−0.056	
	(0.104) [0.480]	(0.095) [0.555]	
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*) 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 1. 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 J.9). 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.

Figure A52: *CDF of the number of times transgender workers are selected in a group shows no evidence of increased polarisation*



Notes: X-axis shows the number of times a transgender worker is selected by a group-of-3 in the private outcome round. Each participant is shown 2 choices with a transgender worker, so the maximum value of this quantity is 6. Y-axis is the cumulative proportion of groups for each value. Sample includes only 3-person discussion and No discussion (private) arms, including both phases 1 and 2.

Table A53: *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 A54: *Effect of listening to a discussion that selects 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							0.0791279245738711	0.105205212455434

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 J.9).

Table A55: *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	−0.001	0.039
	(0.015) [0.998]	(0.014) [0.961]	(0.024) [0.112]
Listener (2-person discussion)	0.012	0.003	0.120***
	(0.020) [0.535]	(0.019) [0.891]	(0.032) [<0.001]
Discussion (pooled)	0.004	0.009	0.148***
	(0.013) [0.730]	(0.012) [0.447]	(0.022) [<0.001]
Worker is trans × Observer (No discussion, public)	0.045*	0.040	
	(0.027) [0.096]	(0.026) [0.126]	
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***	0.144***	
	(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	0.036	0.022
p(Observer=Discussion)	0.000	0.000	0.000
p(Listener=Discussion)	0.430	0.623	0.399

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 1. 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 J.9). 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 A56: *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 J.8 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 J.6 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 rating 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 Estimation of WTP to avoid transgender workers

In this section, I use the generalized method of moments to identify the monetary value of the distaste participants have for selecting a transgender worker in each treatment group. The procedure is similar to the structural estimation carried out in Rao (2019).

I assume that the latent utility received by participant i from group j from choosing worker w in choice k is:

$$U_{ijkw} = \alpha Alternative_{ijkw} + V_{ijkw} - D_{ij}T_{ijkw} + u_{ijkw}$$

where:

- V_{ijk} is the monetary value of the items being offered.
- T_{ijkw} = 1 when the worker w being shown is transgender, and 0 otherwise
- D_{ij} is i's distaste for selecting and interacting with a transgender worker, which is assumed to be constant across different choices that i makes.
- $Alternative_{ijkw} = 1$ when the worker w is the alternative worker. $\alpha > 0$ allows for an innate preference for women that appear as alternative workers.

So the participant will select the alternative worker, denoting the choice as $Y_{ijk} = 1$ for selecting $w = 1$, iff:

$$\begin{aligned} \Delta U_{ijk} &= \alpha + \Delta\ V_{ijk} - D_{ij}T_{ijk} + \Delta u_{ijk} \\ &= \alpha + \Delta\ V_{ijk} - D_{ij}T_{ijk} + \Delta\kappa_j + \Delta\eta_{ij} + \Delta\varepsilon_{ijk} \end{aligned}$$

- ΔV_{ijk} is observed.
- Assume that the preference shock term Δu_{ijk} is split into three components that are all normally distributed: a group-specific term, an individual-specific term, and a choice-specific term:

$$\begin{aligned} \Delta u_{ijk} = \Delta\kappa_j + \Delta\eta_{ij} + \Delta\varepsilon_{ijk}; \\ \Delta\kappa_j \sim \mathcal{N}(0, \sigma_\kappa^2); \\ \Delta\eta_{ij} \sim \mathcal{N}(0, \sigma_\eta^2); \\ \Delta\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{aligned}$$

- For each individual, their distaste can be correlated with the other individuals in their group. So distaste is divided into a component that is common to the group, and a component that is individual-specific:

$$D_{ij} = \bar{D}_j + d_{ij}$$

  The mean and variance of the group-level component is allowed to across treatment groups $g \in \{Discuss, Control\}$ i.e.:

$$\begin{aligned} \bar{D}_j &\sim \mathcal{N}(\mu(g), \sigma_D^2(g)) \\ d_{ij} &\sim \mathcal{N}(0, \sigma_d^2) \end{aligned}$$

So there are 9 parameters to estimate: $[\mu(Discuss), \mu(Control), \sigma_D^2(Discuss), \sigma_D^2(Control), \sigma_d^2, \sigma_\kappa^2, \sigma_\eta^2, \sigma_\varepsilon^2, \alpha]$

B.1 Moments

I use three sets of moments to estimate the model: the mean probabilities of selecting the alternative worker, the intraclass correlation within individuals, and the intraclass correlation within groups. For each of these sets, I calculate the moments for each of the two treatment groups, for transgender and non-transgender alternative workers, and for each value of $\Delta V_{ijk} \in \{-172, -86, 0, 86, 176\}$. There are therefore $3 \times 2 \times 2 \times 5 = 60$ moments in total.

I describe below the theoretical moments predicted by a given set of parameters.

B.1.1 Mean probability of selecting the alternative worker

I define the model-based means of $P(Y_{ijk} = 1|\Delta V_{ijk}, T_{ijk}, g)$, i.e. the means of the outcome variable for each value of ΔV_{ijk}, in each treatment group, for transgender and non-transgender workers.

$$
\begin{aligned}
& P\left[Y_{ijk} = 1|\Delta V_{ijk}, T_{ijk}, g\right] \\
&= P\left[\Delta U_{ijk} \geq 0|\Delta\ V_{ijk}, T_{ijk}, g\right] \\
&= P\left[\alpha + \Delta V_{ijk} - D_{ij}T_{ijk} + \Delta\ u_{ijk} \geq 0|\Delta V_{ijk}, T_{ijk}, g\right] \\
&= P\left[\alpha + \Delta V_{ijk} - (\bar{D}_j + d_{ij})T_{ijk} + \Delta\kappa_j + \Delta\ \eta_{ij} + \Delta\varepsilon_{ijk} \geq 0|\Delta V_{ijk}, T_{ijk}, g\right] \\
&= P\left[\Delta\kappa_j + \Delta\ \eta_{ij} + \Delta\varepsilon_{ijk} - (\bar{D}_j - \mu(g) + d_{ij})T_{ijk} \geq -\alpha - \Delta V_{ijk} + \mu(g)T_{ijk}\right] \\
&= P\left[\frac{\Delta\kappa_j + \Delta\ \eta_{ij} + \Delta\varepsilon_{ijk} - (\bar{D}_j - \mu(g) + d_{ij})T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + \left(\sigma_D^2 + \sigma_d^2\right) T_{ijk}} \geq \frac{-\alpha - \Delta V_{ijk} + \mu(g)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + \left(\sigma_D^2 + \sigma_d^2\right) T_{ijk}}\right] \\
&= \Phi\left(\frac{\alpha + \Delta V_{ijk} - \mu(g)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + \left(\sigma_D^2 + \sigma_d^2\right) T_{ijk}}\right)
\end{aligned}
$$

where $\Phi(.)$ is the CDF of the standard normal distribution.

B.1.2 Intra-class correlations

To identify the variance terms in the model, I use different measures of intra-class correlation as moments.

First, note that the Pearson's correlation coefficient between two binary variables Y_A and Y_B based on a latent index with the same marginal probabilities is (Rodríguez & Elo, 2003):

$$\rho_Y = \frac{\pi_{11} - \pi_{.1}^2}{\pi_{.1}(1 - \pi_{.1})} \tag{3}$$

where $\pi_{.1}$ denotes the marginal probability of $Y_A = Y_B = 1$, and π_{11} denotes the joint probability of both $Y_A = 1$ and $Y_B = 1$.

The marginal probability of $Y_{ijk} = 1$ is:

$$\pi_{.1} = \varphi\left(\frac{\alpha + \Delta V_{ijk} - \mu(g)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + \left(\sigma_D^2 + \sigma_d^2\right) T_{ijk}}\right)$$

where $\varphi(.)$ is the PDF of the standard normal distibution.

The joint probability π_{11} will be dependent on the correlation between the latent indexes. There are two cases.

Case 1: Correlation *within* individuals, i.e. comparing $i = i, j = j, k \neq k'$.

For simplicity, I impose that that $\Delta V_{ijk} = \Delta V_{ijk'}$ and $T_{ijk} = T_{ijk'}$, i.e. I only compare within these cells rather than between them.

First, consider the covariance in the latent utilities:

$$\begin{aligned}
&Cov(\Delta\ U_{ijk}, \Delta U_{ijk'} | \Delta\ V_{ijk}, T_{ijk}, g) \\
&= Cov[\alpha + \Delta\ V_{ijk} - (\bar{D}_j + d_{ij})T_{ijk} + \Delta\ \kappa_j + \Delta\eta_{ij} + \Delta\varepsilon_{ijk}, \\
&\qquad \alpha + \Delta\ V_{ijk'} - (\bar{D}_j + d_{ij})T_{ijk} + \Delta\ \kappa_j + \Delta\eta_{ij} + \Delta\varepsilon_{ijk'}] \\
&= \sigma_\kappa^2 + \sigma_\eta^2 + (\sigma_D^2 + \sigma_d^2)T_{ijk}
\end{aligned}$$

The total variance in the latent utilities is:

$$Var(\Delta U_{ijk}) = \sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + (\sigma_D^2 + \sigma_d^2)T_{ijk}$$

So the Pearson correlation coefficient between the two latent utilities is:

$$\rho(\Delta\ U_{ijk}, \Delta U_{ijk'}) = \frac{\sigma_\kappa^2 + \sigma_\eta^2 + (\sigma_D^2 + \sigma_d^2)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + (\sigma_D^2 + \sigma_d^2)T_{ijk}}$$

So using Equation 3, the correlation between the outcome variables is:

$$\rho_Y(Y_{ijk}, Y_{ijk'}) = \frac{\pi_{11}(\rho(\Delta U_{ijk}, \Delta U_{ijk'})) - \pi_{.1}^2}{\pi_{.1}(1 - \pi_{.1})}$$

where $\pi_{.1}$ is defined as above, and:

$$\begin{aligned}
&\pi_{11}(\rho(\Delta U_{ijk}, \Delta U_{ijk'})) \\
&:= \Phi_2\left(\frac{\alpha + \Delta V_{ijk} - \mu(g)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + (\sigma_D^2 + \sigma_d^2)\ T_{ijk}}, \frac{\alpha + \Delta V_{ijk} - \mu(g)T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + (\sigma_D^2 + \sigma_d^2)\ T_{ijk}}, \rho(\Delta U_{ijk}, \Delta U_{ijk'})\right)
\end{aligned}$$

i.e., the joint probability is based on a standard bivariate normal distribution with a correlation of $\rho(\Delta U_{ijk}, \Delta U_{ijk'})$ between the two variables.

Case 2: Correlation between individuals within a group, i.e. comparing $i \neq i', j = j, k \neq k'$.

Again, for simplicity, only compare within cells, so impose that $\Delta V_{ijk} = \Delta V_{i'jk'}$ and $T_{ijk} = T_{i'jk'}$.

The covariance in latent utilities is now:

$$\begin{aligned}
&Cov(\Delta\ U_{ijk}, \Delta U_{i'jk'} | \Delta\ V_{ijk}, T_{ijk}, g) \\
&= Cov[\alpha + \Delta\ V_{ijk} - (\bar{D}_j + d_{ij})T_{ijk} + \Delta\ \kappa_j + \Delta\eta_{ij} + \Delta\varepsilon_{ijk}, \\
&\qquad \alpha + \Delta\ V_{i'jk'} - (\bar{D}_j + d_{i'j})T_{ijk} + \Delta\ \kappa_j + \Delta\eta_{i'j} + \Delta\varepsilon_{ijk'}] \\
&= \sigma_\kappa^2 + \sigma_D^2 T_{ijk}
\end{aligned}$$

So the correlation coefficient is:

$$\rho(\Delta\ U_{ijk}, \Delta U_{i'jk'}) = \frac{\sigma_\kappa^2 + \sigma_D^2 T_{ijk}}{\sigma_\kappa^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 + (\sigma_D^2 + \sigma_d^2) T_{ijk}}$$

Which allows us to define $\rho_Y(Y_{ijk}, Y_{i'jk'})$ in the same way as Case 1.

B.2 Estimation

To estimate the model, I use the generalized method of moments. I use a minimum distance estimator that solves $\min_\theta (m(\theta) - \hat{m})' W (m(\theta) - \hat{m})$, where θ is the vector of parameters, $m(\theta)$ is the set of theoretically predicted moments based on those parameters, and $\hat{m}$ is the empirically estimated moments from the data. For W, the weighting matrix, I use the diagonalised inverse covariance for each moment. This means that moments that are more precisely estimated in the data receive more weight in the estimation. I solve the equation using numerical optimisation based on a quasi-Newton algorithm that constrains the variances to be weakly greater than 0 (Byrd et al., 1995).

Identification. The exogenous variation in ΔV_{ijk} identifies both mean and the variance of the distaste D in each treatment group, i.e. identifies $\mu(Discuss), \mu(Control), \sigma_D^2(Discuss), \sigma_D^2(Control)$. To identify each of the variance terms, note that: (i) σ_κ^2 is identified by the within-group correlation for non-transgender choices; (ii) σ_D^2 is identified by comparing the latter with the within-group correlation for transgender choices; (iii) σ_η^2 is identified by the additional within-individual correlation for non-transgender choices; (iv) σ_d^2 is identified by the additional within-individual correlation for transgender choices; (v) σ_ε^2 is the remaining unexplained variance.

B.3 Results

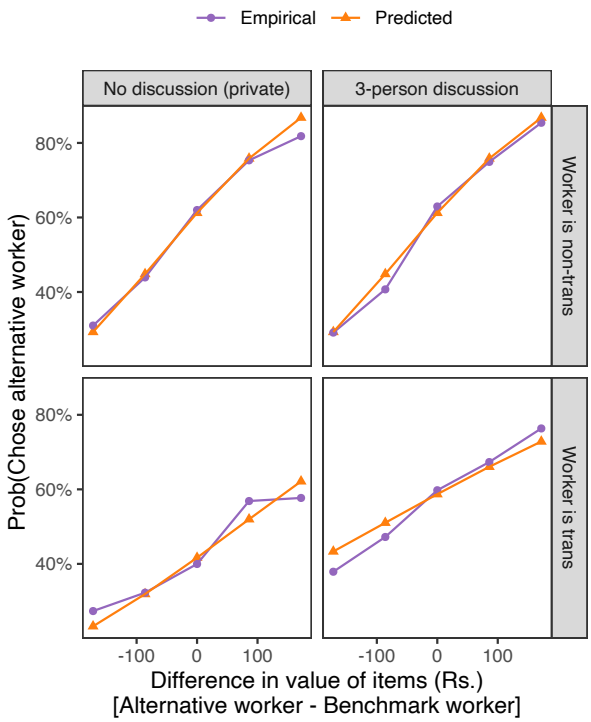
Table B1 shows the estimated parameters. In the control group, I estimate a large distaste for transgender people of 129 Rs. In the 3-person discussion group, this distaste is estimated to be slightly negative at -38 Rs, implying a small positive preference for selecting transgender workers. Figure B2 shows that the model performs well at predicting the empirical probabilities of selecting a transgender worker in each cell.

Table B1: *Structural estimates*

Parameter	Value
$\mu(Discuss)$	-38.4
$\mu(Control)$	128.7
$\sigma_D^2(Discuss)$	123.8
$\sigma_D^2(Control)$	236.4
σ_d^2	0.0
σ_ε^2	85.8
σ_κ^2	56.5
σ_η^2	64.7
α	59.1

Notes: Results from the structural estimation of the model in Section B.

Figure B2: *Structural estimation of WTP to avoid transgender workers: comparison between predicted moments and empirical moments*



Notes: The graph displays both the theoretically predicted and the empirically estimated proportions of people who choose the alternative worker for both transgender and non-transgender workers ($T_{ijk} \in \{0,1\}$), for each treatment group ($g \in \{Discuss, Control\}$), and for each value of $\Delta V_{ijk} \in \{-172, -86, 0, 86, 176\}$.

C Distribution of discrimination: positive and negative discrimination, and polarisation

Positive vs negative discrimination. To shed more light on the changes in discrimination driven by the discussion, I examine here the extent to which effects are driven by a reduction in negative discrimination or an increase in positive discrimination.

To do this, I divide up the hiring choices that participants make into three categories based on the items offered and the reliability score of each pair: (i) *dominates*, in which the alternative worker option is weakly preferable in terms of items and reliability score; (ii) *dominated*, in which the alternative worker option is weakly worse in terms of item and reliability score; and (iii) *neither dominates*, in which neither option dominates in terms of item and reliability score (e.g. they both offer the same items and reliability score is not shown).[44] If a transgender worker is selected when they are dominated, I interpret that as evidence of positive discrimination, i.e., actively favoring a transgender worker in order to help them. If a transgender worker is *not* selected when they dominate, I interpret that as evidence of negative discrimination.

In the outcome round, discussions increase the probability of selecting a transgender uniformly across the three domination categories. Table A15 and Figure A16 show that the interactions between the main treatment effect on (Worker is trans $\times$ Discussion (pooled)) and whether the worker dominates or is dominated have point estimates very close to 0 (with p-values of 0.87 and 0.59, respectively). This implies that discussions simultaneously reduced negative discrimination and increased positive discrimination.

The choices made *during* the discussion also show evidence of a decrease in negative discrimination and an increase in positive discrimination (Table A15, column 2). The increase in positive discrimination is particularly strong, however: the interaction term (Worker is trans $\times$ Discussion (pooled) $\times$ Dominated) is positive and significant (p-value: 0.06), suggesting that during the discussion people are especially likely to positively discriminate towards transgender workers. Indeed, looking at the means in Figure A16 shows that even when faced with a *dominated* transgender worker during a discussion, people select the transgender worker 62% of the time, suggesting significant positive discrimination in favor of transgender workers.

Polarisation. Leading theories of group dynamics suggest that groups often exhibit more extreme behavior than individuals because convergence within a group can lead to more polarisation between groups (Davis, 1973; Kerr et al., 1979; Stasser & Davis, 1981; Kerr, 1981; Ambrus et al., 2015). In the current setting, there is indeed convergence in choices within a group after discussion: the intracluster correlation coefficient based on whether a participant selects a transgender in the outcome round is higher for group discussion participants (Table A35). However, this does not translate to more extreme behavior between groups. In Figure A52, I examine the distribution of how many times a group selects a transgender worker in their individual outcome round choices (out of 6, since there are 3 people in each group, each of whom has 2 opportunities to select a transgender worker). If

[44] See the notes of Table A15 for the precise definitions used.

group discussions increased polarisation, this would lead to increased clustering around the poles of 0 and 6. But we see that in fact the distribution for *3-person discussion* first-order stochastically dominates the distribution for *No discussion (private)*. In other words, the number of transgender workers selected in the outcome round increases across the whole distribution.

D Ethical considerations

Understanding how to reduce discrimination towards the transgender community in India is of great social importance. The study was designed to yield a revealed-preference measure of discrimination based on choices with real stakes. Having real choices rather than hypothetical choices was crucial for measuring discrimination in a way that was less vulnerable to concerns about social desirability bias and experimenter demand effects, and therefore for understanding methods for reducing such discrimination. However, it meant that the design had to trade off multiple ethical considerations - most notably trying to avoid explicitly deceiving respondents while also protecting transgender workers from harm.

There is an important concern that the protocol may be seen as deceptive for the participants because they were unlikely to receive a delivery from a transgender worker. This concern, nevertheless, had to be balanced against the risks that transgender workers would have faced if they had carried out deliveries. If transgender workers were to visit the homes of participants, they could have been vulnerable to stigma and verbal or physical abuse. The randomization was designed to avoid this as much as possible while also truthfully telling participants that they could receive a delivery from any of the workers they chose. For the few transgender workers who actually carried out a delivery, the transgender worker was accompanied by a full team of 2-3 enumerators throughout the entire process. Interaction between the transgender worker and participant was reduced to a minimum, and the other enumerators were extensively trained to avoid conflict and protect the worker. Strict protocols were put in place to ensure that the survey-team member was not confused with the transgender worker and, thus, was not put at risk themselves. This design protected the transgender workers as much as possible while also not deceiving the respondents and being able to elicit truthful revealed preference responses.

A second concern is that while the reliability score shown on some of the worker profiles was truthful, participants were not given enough context to interpret it correctly. Using the reliability score was important for examining whether discrimination against transgender workers in this context was mostly *taste-based* or *statistical*. If anti-transgender discrimination was reduced when the score was displayed, this would indicate that statistical discrimination contributed to the total discrimination they faced. This distinction has important policy implications, since it may be easier to reduce statistical discrimination by informing about the productivity of transgender workers, whereas reducing taste-based discrimination requires changing private attitudes or leveraging social pressure (as in the current paper).

The reliability score was a truthful report of how many deliveries a worker successfully carried out in a training exercise. However, participants were unaware that workers carried out multiple training exercises with different durations, and that these yielded different scores. Participants were told (when shown an example with a score of 8/10): *"8/10 means that out of 10 deliveries they had to make in a training exercise, 8 times they successfully delivered."* Importantly, everything they were told was true, and they were not given any additional false detail about the nature of the exercise: for example, they were not told that the exercise was only done

once. This mimics real-world situations, in which employers often have incomplete sources of information about job candidates.

Finally, the primary reason participants cared about the reliability of the worker was to ensure that they would actually receive a delivery. In practice, we completed deliveries to 95.7% of the participants. This implies that participants were always in fact choosing a delivery worker who was extremely reliable, and so any misleading inference from the reliability score did not cost them materially.

E Background: Legal context

The legal landscape surrounding the rights of transgender individuals in India has seen significant developments in the past decade, culminating in several landmark judgments and legislative actions.

Central to these advances was the landmark judgment in the case of National Legal Services Authority vs. Union of India and others [(2014) 5 SCC 438] (also called the NALSA judgement). This judgement paved the way for the broader recognition and protection of transgender rights in India. It served served as a cornerstone in recognizing the rights of transgender persons in India. Key directives and observations from the judgment included:

1. *Self-Identification*: The court upheld the right of transgender persons to self-identify their gender as male, female, or third gender. This was a significant recognition of individual autonomy and personal integrity.

2. *Safeguarding Rights*: The court directed the government to take steps to safeguard the rights of transgender persons, including provisions for reservations in educational institutions and public appointments.

3. *Legal Recognition*: Transgender persons were granted legal recognition, and the Court mandated the government to grant them all fundamental rights and equal opportunities by law.

4. *Welfare Measures*: The court directed the state to devise social welfare schemes for the betterment of transgender individuals, focusing on their healthcare, education, and socio-economic conditions.

In response to the directives issued by the Supreme Court in the NALSA judgment, the Indian government initiated legislative action to secure the rights and welfare of transgender individuals, culminating in the enactment of the Transgender Persons (Protection of Rights) Act in 2019. The Act aimed to provide a formal structure to the rights, protections, and entitlements of transgender persons in India. The main provisions of this act were:

1. *Definition of Transgender Persons*: The Act sought to define transgender persons as individuals whose gender does not match the one assigned at birth. It includes trans-men, trans-women, persons with intersex variations, and genderqueers.

2. *Prohibition of Discrimination*: The Act prohibits discrimination against transgender persons in various sectors including education, employment, and healthcare.

3. *Recognition of Identity*: The Act provides for the legal recognition of the transgender person's self-perceived gender identity through a certificate of identity.

4. *Welfare Measures*: The Act mandates the government to formulate welfare measures and policies to secure full participation of transgender individuals in society.

5. *National Council for Transgender Persons*: The Act stipulated the establishment of a National

Council for Transgender persons to advise the government on policies and legislations related to transgender persons.

However, this Act was the subject of a number of protests and controversies, in particular because (i) it did not allow self-identification, in contrast to the NALSA judgement; (ii) it did not allow for affirmative action quotas for transgender persons; and (iii) the Act prescribed lesser penalties for sexual violence against transgender persons compared to cisgender women, a provision which many claimed reinforced the discriminatory stance towards transgender individuals. Because the Act is of disputed merit, the expressive power of this law to alleviate discriminatory norms may have been diluted, so I use the NALSA judgement as the focus of the video about transgender rights in the experiment.

F Proofs

F.1 Proposition 1

I restrict the set of equilibria to equilibria in which everyone choosing $Y_i = 1$ chooses the same S_i and everyone choosing $Y_i = 0$ chooses the same S_i. This means that in a candidate equilibrium, participants only have two choices. To define those two choices, first let $S_i(Y_i; Q)$ be the message sent in a candidate equilibrium by the participants who choose Y_i, for a given candidate equilibrium $Q \in \{SS, SN, NS, NN\}$. Participants can therefore choose either $(Y_i = 1, S_i = S_i(1; Q))$ or $(Y_i = 1, S_i = S_i(0; Q))$.

The four candidate equilibria are defined by how participants' choices of S_i map onto their choices of Y_i:

$$\begin{aligned} S_i(1; SS) &= S_i(0; SS) = 1 \\ S_i(1; SN) = 1, \quad & S_i(0; NS) = 0 \\ S_i(1; NS) = 0, \quad & S_i(0; NS) = 1 \\ S_i(1; NN) &= S_i(1; NN) = 0 \end{aligned}$$

As in Bénabou & Tirole (2006), since the action space in a candidate equilibrium has only two options, the candidate equilibrium is defined by a single threshold. The marginal agent defines the threshold, where the marginal agent has $P_i = P_Q^*$. All agents with $P_i > P_Q^*$ choose $Y_i = 1$ and $S_i(1, Q)$, and all agents with $P_i \leq P_Q^*$ choose $Y_i = 0$ and $S_i(0, Q)$.

Given the utility function in 2, participant i will choose $(1, S_i(1, Q))$ instead of $(0, S_i(0, Q))$ when:

$$\Delta U_i(P_i, Q) := \mathbb{E}_i[U_i(1, S_i(1, Q)|P_i, Q))] - \mathbb{E}_i[U_i(0, S_i(1, Q)|P_i, Q))] \geq 0$$

where I replace σ_{-i} by Q, since σ_{-i} is fixed by the equilibrium type Q.

Writing out $\Delta U_i(P_i, Q)$ in full yields:

$$\begin{aligned} \Delta U_i(P_i, Q) = \Delta V + P_i + \gamma_0 & \left[\mathbb{E}_{-i}[P_i|1, S_i(1; Q), Q] - \mathbb{E}_{-i}[P_i|0, S_i(1; Q), Q]\right] \\ & - \gamma_1 \Big(\mathbb{E}_{-i}[P_i|1, S_i(1; Q), Q] - \mathbb{E}_i[P_{-i}|S_i(1; Q), Q] \\ & + \gamma_1 \Big(\mathbb{E}_{-i}[P_i|0, S_i(0; Q), Q] - \mathbb{E}_i[P_{-i}|S_i(0; Q), Q] \Big)^2 \\ & - c \cdot \Big(|S_i(1; Q)| - |S_i(0; Q)| \Big) \end{aligned} \tag{4}$$

where $\Delta V := V(Y_i = 1) - V(Y_i = 0)$ is the difference between the value of items across the pair. In the following sections, I derive expressions for each of the expectations in terms of the parameters of the model.

F.1.1 Expected P_{-i}

Let π_1 be the proportion of people who choose $Y_i = 1$. Given a cutoff point P_Q^*, and using the uniform distribution of P_i, this will be equal to:

$$\pi_1(P_Q^*) = \begin{cases} \frac{\mu_P + R - P_Q^*}{2R} & \text{if } P_Q^* \in [\mu_P - R, \mu_P + R] \\ 1 & \text{if } P_Q^* < \mu_P - R \\ 0 & \text{if } P_Q^* > \mu_P + R \end{cases}$$

Now, I derive the expressions for $\mathbb{E}_i[P_{-i}|S_i(Y_i;Q), Q]$, i.e., for i's expectation of the value of other people's P_{-i} in each type of equilibrium, after accounting for any persuasion that occurs. I use the fact that in each type of equilibrium, people's choice of Y_i will also determine their choice of S_i.

1. **SS equilibrium**:

$$\begin{aligned} \mathbb{E}_i[P_{-i}|S_i, Q = SS] &= \mu_P + \pi_1\alpha - (1-\pi_1)\alpha + \alpha S_i \\ &= \begin{cases} \mu_P + \frac{\alpha(\mu_P - P_{SS}^*)}{R} + \alpha S_i & \text{if } P_{SS}^* \in [\mu_P - R, \mu_P + R] \\ \mu_P + \alpha + \alpha S_i & \text{if } P_{SS}^* < \mu_P - R \\ \mu_P - \alpha + \alpha S_i & \text{if } P_{SS}^* > \mu_P + R \end{cases} \end{aligned}$$

2. **SN equilibrium**:

$$\begin{aligned} \mathbb{E}_i[P_{-i}|S_i, Q = SN] &= \mu_P + \pi_1\alpha + \alpha S_i \\ &= \begin{cases} \mu_P + \frac{\alpha(\mu_P + R - P_{SN}^*)}{2R} + \alpha S_i & \text{if } P_{SN}^* \in [\mu_P - R, \mu_P + R] \\ \mu_P + \alpha + \alpha S_i & \text{if } P_{SN}^* < \mu_P - R \\ \mu_P + \alpha S_i & \text{if } P_{SN}^* > \mu_P + R \end{cases} \end{aligned}$$

3. **NS equilibrium**:

$$\mathbb{E}_i[P_{-i}|S_i, Q = NS] = \begin{cases} \mu_P + \frac{\alpha(P_{NS}^* - \mu_P + R)}{2R} + \alpha S_i & \text{if } P_{NS}^* \in [\mu_P - R, \mu_P + R] \\ \mu_P + \alpha S_i & \text{if } P_{NS}^* < \mu_P - R \\ \mu_P - \alpha + \alpha S_i & \text{if } P_{NS}^* > \mu_P + R \end{cases}$$

4. **NN equilibrium**:

$$\mathbb{E}_i[P_{-i}|S_i, Q = NN] = \mu_P + \alpha S_i$$

F.1.2 Expected P_i

Here, I derive the expressions for $\mathbb{E}_{-i}[P_i|Y_i, S_i(Y_i;Q), Q]$, i.e., i's expectation of what other's expectation of her type P_i will be if she takes action $(Y_i, S_i(Y_i;Q))$. I take into account expected persuasion that occurs in equilibrium. i is not persuaded to change her P_i by her *own* message S_i, but correctly anticipates the average amount of persuasion due to other people's messages.

This means that relevant distribution of P_i to be used is centred at $\mathbb{E}_i[P_{-i}|S_i = 0, Q]$ as defined above, since this expression accounts for persuasion coming from other group members, but not i herself.

In equilibrium, all participants with $P_i > P_Q^*$ choose $Y_i = 1$. This means that the expected P_i having been observed to choose $Y_i = 1$ will be the average value of P_i for those above P_Q^*.

Focusing on interior solutions for P_Q^* first, and using the uniform distribution of P_i, the expectation of P_i if i chooses $Y_i = 1$ is therefore:

$$\mathbb{E}_{-i}[P_i|Y_i = 1, S_i(1;Q), Q] = \frac{P_Q^* + \mathbb{E}_i[P_{-i}|S_i = 0, Q] + R}{2} \qquad \text{if } P_Q^* \in [\mu_P - R, \mu_P + R]$$

Similarly, the conditional expectation when choosing $Y_i = 0$ is:

$$\mathbb{E}_{-i}[P_i|Y_i = 0, S_i(0;Q), Q] = \frac{P_Q^* + \mathbb{E}_i[P_{-i}|S_i = 0, Q] - R}{2} \qquad \text{if } P_Q^* \in [\mu_P - R, \mu_P + R]$$

Corner solutions:

Corner solutions are where either *everyone* chooses $Y_i = 1$ (when $P_Q^* < \mu_P - R$), or *everyone* chooses $Y_i = 0$ (when $P_Q^* > \mu_P + R$). In such cases, I assume that off-equilibrium beliefs are as follows:

1. When everyone chooses $Y_i = 1$, assume that if i chooses $Y_i = 0$ then people believe i to have the *minimum* P_i possible, then:

$$\begin{aligned}\mathbb{E}_{-i}[P_i|Y_i = 1, S_i(1;Q), Q] &= \mu_P + \alpha S_i(1;Q)\\ \mathbb{E}_{-i}[P_i|Y_i = 0, S_i(0;Q), Q] &= \mu_P - R + \alpha S_i(1;Q)\end{aligned}$$

2. When everyone chooses $Y_i = 0$, assume that if i chooses $Y_i = 1$ then people believe that i has the maximum P_i possible, then e.g.:

$$\begin{aligned}\mathbb{E}_{-i}[P_i|Y_i = 1, S_i(1;Q), Q] &= \mu_P + R + \alpha S_i(0;Q)\\ \mathbb{E}_{-i}[P_i|Y_i = 0, S_i(0;Q), Q] &= \mu_P - R + \alpha S_i(0;Q)\end{aligned}$$

F.1.3 Equilibrium conditions

The equilibrium is defined by the marginal agent, for whom we have a fixed point equation, i.e., $\Delta U_i(P_Q^*; Q) = 0$.

Using the expressions from the previous two subsections, and plugging them into equation 4 yields the full equilibrium conditions for each type of equilibrium. These can be rearranged to

derive the value of P^*_Q for each Q:

$$
\begin{aligned}
P^*_{SS} &= \frac{2\alpha^2\gamma_1\mu_P + \alpha\gamma_1\mu_P R - \gamma_0 R^2 - \gamma_1\mu_P R^2 - R\Delta V}{2\alpha^2\gamma_1 + R + \alpha\gamma_1 R - \gamma_1 R^2} \\
P^*_{SN} &= \frac{\alpha\gamma_1(\mu_P - 3R)R + \alpha^2\gamma_1(\mu_P + 3R) - 2R(-c + \gamma_0 R + \gamma_1\mu_P R + \Delta V)}{\alpha^2\gamma_1 + \alpha\gamma_1 R - 2R(-1 + \gamma_1 R)} \\
P^*_{NS} &= \frac{c + \alpha^2\gamma_1 + \gamma_0 R + \gamma_1\mu_P R - \alpha\gamma_1(\mu_P + R) + \Delta V}{\gamma_1(R - \alpha) - 1} \\
P^*_{NN} &= \frac{\gamma_0 R + \gamma_1\mu_P R + \Delta V}{-1 + \gamma_1 R}
\end{aligned}
\tag{5}
$$

F.1.4 Incentive compatibility constraints for sending behavior

The method of calculating P^*_Q ensures incentive compatibility for the choice of Y_i. But to ensure that behavior is an equilibrium, I need to examine the incentive compatibility constraints for each type of player and derive the conditions under which the players have an incentive to send in the way they do under equilibrium.

1. **SS equilibrium**

   For the SS equilibrium, all the following derivations require the assumptions that $\alpha < (R/2)$ and $\gamma_1 < R/(R^2 - \alpha R - 2\alpha^2)$.

   (a) *Corner solutions*: There is a corner solution where everyone chooses $Y_i = 1$ if $P^*_{SS} < \mu_P - R$. An interior solution therefore requires that $\mu_P \leq \eta^{SS}_{Y=1}$ where:

   $$\eta^{SS}_{Y=1} := 2\alpha^2\gamma_1 + R - \gamma_0 R + \alpha\gamma_1 R - \gamma_1 R^2 - \Delta V$$

   There is a corner solution where everyone chooses $Y_i = 0$ if $P^*_{SS} > \mu_P + R$. Under the same assumptions, an interior solution also requires that $\mu_P \geq \eta^{SS}_{Y=0}$, where:

   $$\eta^{SS}_{Y=0} := -2\alpha^2\gamma_1 - (1 + \gamma_0)R - \alpha\gamma_1 R + \gamma_1 R^2 - \Delta V$$

   (b) *Choosers* ($Y_i = 1$):

   For the people who choose $Y_i = 1$, sending $S_i = 1$ has to be weakly preferable to sending $S_i = 0$. So:

   $$E_i[U_i(Y_i = 1, S_i = 1|P_i, SS)] - E_i[U_i(Y_i = 1, S_i = 0|P_i, SS)] \geq 0$$

   Plugging in the values of P^*_{SS} from equation 5 and the expectations from Section F.1.1 yields the result that if $\alpha < (R/2)$ and $\gamma_1 < \frac{R}{R^2 - \alpha R - 2\alpha^2}$, then the equation will hold if $\mu_P \leq \kappa^{SS}_{Y=1}$, where

   $$\kappa^{SS}_{Y=1,S\neq 0} := -\left[\frac{-2\alpha^2\gamma_1 R + \alpha(\gamma_0 R - \gamma_1 R^2 + \Delta V) + R((-1 + \gamma_0)R + \gamma_1 R^2 + \Delta V)}{\alpha + R}\right]$$

   For the people who choose $Y_i = 1$, sending $S_i = 1$ also has to be weakly preferable to sending $S_i = -1$. Under the same assumptions that $\alpha < (R/2)$ and $\gamma_1 < \frac{R}{R^2 - \alpha R - 2\alpha^2}$,

this generates the constraint that $\mu_P \leq \tau^{SS}_{Y=1}$, where:

$$\kappa^{SS}_{Y=1,S\neq-1} := \frac{-2\alpha^4\gamma_1^2 + \alpha^3\gamma_1^2 R + cR(-1+\gamma_1 R) - \alpha^2\gamma_1(2c + R + \gamma_0 R - 2\gamma_1 R^2 + \Delta V) - \alpha\gamma_1 R(c + (-1+\gamma_0)R + \gamma_1 R^2 + \Delta V)}{\alpha\gamma_1(\alpha + R)}$$

For $\alpha < (R/2)$, and $\gamma_1 < \frac{R}{R^2 - \alpha R - 2\alpha^2}$, it is always true that $\kappa^{SS}_{Y=1,S\neq0} < \kappa^{SS}_{Y=1,S\neq-1}$, so only the constraint based on $\kappa^{SS}_{Y=1,S\neq0}$ is binding.

(c) *Non choosers* ($Y_i = 0$):

For people who choose $Y_i = 0$, sending $S_i = -1$ has to be weakly preferable to sending $S_i = 0$.

This yields a constraint $\mu_P \geq \kappa^{SS}_{Y=0,S\neq0}$ where:

$$\kappa^{SS}_{Y=0,S\neq0} := \frac{2\alpha^4\gamma_1^2 - \alpha^3\gamma_1^2 R + cR(1-\gamma_1 R) + \alpha\gamma_1 R(c - (1+\gamma_0)R + \gamma_1 R^2 - \Delta V) - \alpha^2\gamma_1(-2c + (1+\gamma_0)R + 2\gamma_1 R^2 + \Delta V)}{\alpha\gamma_1(\alpha + R)}$$

In addition, $S_i = -1$ has to be weakly preferable to $S_i = 1$, which yields a constraint $\mu_P \geq \kappa^{SS}_{Y=0,S\neq1}$, where

$$\kappa^{SS}_{Y=0,S\neq1} := -\frac{2\alpha^2\gamma_1 R + R(R + \gamma_0 R - \gamma_1 R^2 + \Delta V) + \alpha(\gamma_0 R + \gamma_1 R^2 + \Delta V)}{\alpha + R}$$

For $c > 0$, $\alpha > 0$, $R > 0$, $\alpha < (R/2)$, and $\gamma_1 < \frac{R}{R^2 - \alpha R - 2\alpha^2}$, it is always true that $\kappa^{SS}_{Y=0,S\neq0} > \kappa^{SS}_{Y=0,S\neq1}$, so only the constraint based on $\kappa^{SS}_{Y=0,S\neq0}$ is binding.

A SS equilibrium is feasible when there is some range of values μ_P that satisfies all the constraints outlined above.

To get an interior solution, we require:

$$\mu_P \in [\eta^{SS}_{Y=0}, \eta^{SS}_{Y=1}]$$

and:

$$\mu_P \in [\kappa^{SS}_{Y=0,S\neq0}, \kappa^{SS}_{Y=1,S\neq0}]$$

If $c > -2\alpha^2\gamma_1$ (which always holds by assumption that $c > 0$), we find that $\kappa^{SS}_{Y=1,S\neq0} < \eta^{SS}_{Y=1}$ and $\eta^{SS}_{Y=0} < \kappa^{SS}_{Y=0,S\neq0}$. This implies that the corner condition constraints are not binding.

The SS equilibrium is therefore feasible when:

$$\kappa^{SS}_{Y=0,S\neq0} < \kappa^{SS}_{Y=1,S\neq0}$$

This holds when the cost of communicating is sufficiently low:

$$c < c^* := \alpha\gamma_1 R - \alpha^2\gamma_1$$

i.e. as long as this condition holds, there will be some values of μ_P, i.e., $\mu_P \in [\kappa^{SS}_{Y=0,S\neq0}, \kappa^{SS}_{Y=1,S\neq0}]$, for which SS is an equilibrium.

2. **SN equilibrium**

The derivations for the SN equilibrium require assuming $\alpha < R$ and $\gamma_1 < 2R/(2R^2 - \alpha R - \alpha^2)$.

(a) *Corner solutions*.

To get an interior solution, we require $P^*_{SN} \in [\mu_P - R, \mu_P + R]$, which simplifies to the constraint $\mu_P \in [\eta^{SN}_{Y=0}, \eta^{SN}_{Y=1}]$, where:

$$\eta^{SN}_{Y=1} := c + 2\alpha^2\gamma_1 + R - \gamma_0 R - \alpha\gamma_1 R - \gamma_1 R^2 - \Delta V$$
$$\eta^{SN}_{Y=0} := c + \alpha^2\gamma_1 - R - \gamma_0 R - 2\alpha\gamma_1 R + \gamma_1 R^2 - \Delta V$$

(b) *Choosers ($Y_i = 1$)*:

$S_i = 1$ has to be weakly preferable to $S_i = 0$, which holds when $\mu_P \leq \kappa^{SN}_{Y=1,S\neq 0}$, where:

$$\kappa^{SN}_{Y=1,S\neq 0} := \frac{cR\left(-2 + \alpha\gamma_1 + 2\gamma_1 R\right) + \alpha\gamma_1\left(\alpha^2\gamma_1 R - \alpha\left((3+\gamma_0)R - \gamma_1 R^2 + \Delta V\right) - 2R\left((-1+\gamma_0)R + \gamma_1 R^2 + \Delta V\right)\right)}{\alpha\gamma_1(\alpha + 2R)}$$

As above, the constraint imposed by preferring $S_i = 1$ to $S_i = -1$ is not binding.

(c) *Non-choosers ($Y_i = 0$)*:

$S_i = 0$ has to be weakly preferable to $S_i = -1$, which yields the constraint $\mu_P \leq \kappa^{SN}_{Y=0,S\neq -1}$ where

$$\kappa^{SN}_{Y=0,S\neq -1} := \frac{2\alpha^4\gamma_1^2 + \alpha^3\gamma_1^2 R - 2cR(-1+\gamma_1 R) - \alpha^2\gamma_1(-2c + (-1+\gamma_0)R + 5\gamma_1 R^2 + \Delta V) + \alpha\gamma_1 R(3c - 2(R + \gamma_0 R - \gamma_1 R^2 + \Delta V))}{\alpha\gamma_1(\alpha + 2R)}$$

$S_i = 0$ has to be weakly preferable to $S_i = 1$, which yields the constraint $\mu_P \geq \kappa^{SN}_{Y=0,S\neq 1}$, where:

$$\kappa^{SN}_{Y=0,S\neq 1} := \frac{cR(-2 + \gamma_1 + 2\gamma_1 R - \alpha\gamma_1(\alpha^2\gamma_1 R + 2R(R + \gamma_0 R - \gamma_1 R^2 + \Delta V) + \alpha(3+\gamma_0)R + \gamma_1 R^2 + \Delta V))}{\alpha\gamma_1(\alpha + 2R)}$$

Under the assumptions on α and γ_1, and assuming $c < c^*$, we get: $\kappa^{SN}_{Y=0,S\neq 1} < \eta^{SN}_{Y=0}$, and $\kappa^{SN}_{Y=0,S\neq -1} < \kappa^{SN}_{Y=1,S\neq 0}$, and $\eta^{SN}_{Y=0} < \kappa^{SN}_{Y=0,S\neq -1}$, implying that the set of constraints can be reduced to:

$$\mu_P \in [\eta^{SN}_{Y=0}, \kappa^{SN}_{Y=0,S\neq -1}]$$

3. **NS equilibrium**.

The NS derivations require the assumptions that $\alpha < R$ and $\gamma_1 < 1/(R - \alpha)$.

(a) *Corner solutions*:

To get an interior solution, we require $P^*_{SN} \in [\mu_P - R, \mu_P + R]$, which simplifies to the constraint $\mu_P \in [\eta^{NS}_{Y=0}, \eta^{NS}_{Y=1}]$, where:

$$\eta^{NS}_{Y=1} := -c - \alpha^2\gamma_1 + R - \gamma_0 R + 2\alpha\gamma_1 R - \gamma_1 R^2 - \Delta V$$
$$\eta^{NS}_{Y=0} := -c - \alpha^2\gamma_1 - R\gamma_0 R + \gamma_1 R^2 - \Delta V$$

(b) *Choosers ($Y_i = 1$)*:

$S_i = 0$ has to be weakly preferable to $S_i = 1$, which yields a constraint $\mu_P \geq \kappa^{NS}_{Y=1,S\neq 1}$ where:

$$\kappa^{NS}_{Y=1,S\neq 1} := \frac{-2\alpha^4\gamma_1^2 + 7\alpha^3\gamma_1^2 R - 2cR(-1+\gamma_1 R) - \alpha^2\gamma_1(2c + (-3+\gamma_0)R + 7\gamma_1 R^2 + \Delta V) + \alpha\gamma_1 R(5c + 2((-1+\gamma_0)R + \gamma_1 R^2 + \Delta V))}{\alpha\gamma_1(\alpha - 2R)}$$

$S_i = 0$ must also be weakly preferable to $S_i = -1$, which yields $\mu_P \leq \kappa^{NS}_{Y=1,S\neq -1}$ where:

$$\kappa^{NS}_{Y=1,S\neq -1} := \frac{cR(-2-\alpha\gamma_1 + 2\gamma_1 R) + \alpha\gamma_1\left(\alpha^2\gamma_1 R + 2R(-1+\gamma_0)R + \gamma_1 R^2 + \Delta V - \alpha(R + \gamma_0 R + 3\gamma_1 R^2 + \Delta V)\right)}{\alpha\gamma_1(\alpha - 2R)}$$

As long as $\alpha < R$ and $\gamma_1 < 1/(R-\alpha)$, it is true that $\kappa^{NS}_{Y=1,S\neq 1} \leq \kappa^{NS}_{Y=1,S\neq -1}$, so there is a range of values for μ_P that makes the equilibrium feasible.

(c) *Non choosers ($Y_i = 0$)*:

$S_i = -1$ has to be weakly preferable to $S_i = 0$, yielding a constraint $\mu_P \geq \kappa^{NS}_{Y=0,S\neq 0}$, where

$$\kappa^{NS}_{Y=0,S\neq 0} := -((1+\gamma_0+\alpha\gamma_1)R) + \gamma_1 R^2 - \frac{cR(2+\alpha\gamma_1 - 2\gamma_1 R)}{\alpha\gamma_1(\alpha - 2R)} - \Delta V$$

$S_i = -1$ has to be weakly preferable to $S_i = 1$, but as before, this constraint is not binding.

Under the assumptions on α and γ_1, along with $c < c^*$, we get: $\kappa^{NS}_{Y=0,S\neq 0} < \kappa^{NS}_{Y=1,S\neq 1}$, and $\eta^{NS}_{Y=1} < \kappa^{NS}_{Y=1,S\neq -1}$ so the constraints are:

$$\mu_P \in [\kappa^{NS}_{Y=1,S\neq 1}, \eta^{NS}_{Y=1}]$$

4. **NN equilibrium**.

These derivations require the assumption that $\alpha < R$ and $\gamma_1 < 1/R$.

(a) *Choosers: $Y_i = 1$*

The incentive compatibility constraints yield $c \geq c_1$ and $c \geq c_2$ where:

$$c_1 := -\frac{\alpha\gamma_1\left(\mu_P - R + \gamma_0 R + \gamma_1 R^2 + \alpha(-1+\gamma_1 R) + \Delta V\right)}{-1+\gamma_1 R}$$
$$c_2 := -\frac{\alpha\gamma_1(-\mu_P + R - \gamma_0 R - \gamma_1 R^2 + \alpha(-1+\gamma_1 R) - \Delta V)}{-1+\gamma_1 R}$$

(b) *Non-choosers: $Y_i = 0$*

The incentive compatibility constraints yield $c \geq c_3$ and $c \geq c_4$, where:

$$c_3 := -\frac{\alpha\gamma_1\left(-\mu_P - R - \gamma_0 R + \gamma_1 R^2 + \alpha(-1+\gamma_1 R) - \Delta V\right)}{-1+\gamma_1 R}$$
$$c_4 := -\frac{\alpha\gamma_1\left(\mu_P + R + \gamma_0 R - \gamma_1 R^2 + \alpha(-1+\gamma_1 R) + \Delta V\right)}{-1+\gamma_1 R}$$

This implies that if $c \geq \bar{c} := \max\{c_1, c_2, c_3, c_4\}$, then the NN equilibrium will be feasible.

F.2 Proposition A1

I consider the case with no persuasion, and show that making choices public can have a null effect on the probability of selecting a transgender worker, even when participants care about virtue signaling ($\gamma_0 > 0$) and about conformity ($\gamma_1 > 0$).

Proposition A1. *Let $y := E(Y_i)$ be the mean probability of selecting a transgender worker in a group. If no persuasion is possible, i.e. $S_i = 0$ for all i, then there is a value of μ_P, call it $\tilde{\mu}_P$, that equalises the virtue signaling and conformity forces, such that $y(\gamma_0, \gamma_1) = y(0, 0)$ with $\gamma_0 > 0$ and $\gamma_1 > 0$.*

The intuition here is that when μ_P is negative, conformity discourages selecting a transgender worker, while virtue signaling encourages it. These forces can balance out at a critical value $\tilde{\mu}_P$.

The probability of selecting a transgender worker y is given by:

$$\begin{aligned} y &= \Pr(Y_i = 1) \\ &= \Pr\left(E\left[U_i(Y_i = 1, S_i = 0)\right] - E\left[U_i(Y_i = 0, S_i = 0)\right] > 0\right) \\ &= \Pr\left(P_i > P^*_{NN}(\gamma_0, \gamma_1)\right) \end{aligned}$$

where $P^*_{NN}(\gamma_0, \gamma_1)$ is the value of P_i for the marginal agent.

This means that $y(\gamma_0, \gamma_1) = y(0, 0)$ if and only if $P^*_{NN}(\gamma_0, \gamma_1) = P^*_{NN}(0, 0)$

First, consider the case where $\gamma_0 > 0$ and $\gamma_1 > 0$, with no persuasion $S_i = 0$ for all i.

In this case, the marginal agent is defined by the fixed point equation:

$$\Delta V + P^*_{NN} + \gamma_0 \left[\mathcal{M}^+(P^*_{NN}) - \mathcal{M}^-(P^*_{NN})\right] - \gamma_1 \left[(\mathcal{M}^+(P^*_{NN}) - \mu_P)^2 - (\mu_P - \mathcal{M}^-(P^*_{NN}))^2\right] = 0$$

where $\Delta V := V(Y_i = 1) - V(Y_i = 0)$ is the difference between the value of items across the pair, and

$$\begin{aligned} \mathcal{M}^+(p) &:= E[P_i | P_i > p] \\ \mathcal{M}^-(p) &:= E[P_i | P_i < p] \end{aligned}$$

are the conditional expectations of P_i above and below some cutoff p. Given the known uniform distribution distribution it is drawn from, $P_i \sim \text{Unif}[\mu_P - R, \mu_P + R]$, we can write:

$$\begin{aligned} \mathcal{M}^+(p) &= \frac{\mu_P + R + p}{2}; \\ \mathcal{M}^-(p) &= \frac{\mu_P - R + p}{2} \end{aligned}$$

Using the values of these conditional expectations for the uniform distribution, we can derive an expression for P^*_{NN}:

$$P^*_{NN}(\gamma_0, \gamma_1) = -\left[\frac{\gamma_0 R + \gamma_1 \mu_P R + \Delta V}{1 - \gamma_1 R}\right]$$

So:

$$
\begin{aligned}
& P^*_{NN}(\gamma_0, \gamma_1) = P^*_{NN}(0,0) \\
\iff & \left[\frac{\gamma_0 R + \gamma_1 \mu_P R + \Delta V}{1 - \gamma_1 R}\right] = \Delta V \\
\iff & \mu_P = \tilde{\mu}_P := -\left[\frac{\gamma_0 R + 2\Delta\ V - \gamma_1 R \Delta\ V}{\gamma_1 R}\right]
\end{aligned}
$$

Therefore $y(\gamma_0, \gamma_1) = y(0,0)$ with $\gamma_0 > 0$ and $\gamma_1 > 0$ if and only if $\mu_P = \tilde{\mu}_P$.

G 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.

G.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 would be 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.

G.2 Updates for phase 2

In phase 2, I made the following changes to the design:

- 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.

G.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 transcribe the discussion recordings and encode a "Probability of endorsing" variable from these transcripts. However, the budget remaining at the end of the experiment did not allow for this, and enumerator observations had already captured this data, so I do not include this for analysis.

G.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 full interacted specification that includes all video arms and all the 3-person discussion arm variation was pre-specified, and is shown in Figure A13 and Table A12.

- *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 (Table A34). However, for reader clarity in the main text I did not pool any treatment arms when presenting the phase 2 outcome round results (Table 5). The corresponding pooled results are presented in Table A55.

- *Heterogeneity with respect to round 1 observations*. Table A53 shows the heterogeneous effects of *observers*'s choices with respect to the round 1 choices they observed. Table A54 shows the heterogeneous effects of *listener*'s choices with respect to the round 1 choices they listened to.

- *Heterogeneity with respect to group composition*. Table A56 shows heterogeneous effects of the discussion with respect to persuasiveness and group relations. I find no detectable heterogeneity.

H 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 H.1).

H.1 Written discussion script for facilitator

1. Explain discussion
We will now have a group discussion. We value your opinion a lot. So for this discussion:
- We want to hear your opinion.
- Each person's experiences and opinions are important.
- If you don't agree, you will need to convince the other people in your group.

1b. Consent to audio recording
To make sure we fully understand your discussion, we would like to record the audio. The audio will be fully anonymized, no-one will know that it was you that was talking in the audio. Do you agree to us recording the audio?

2. Icebreaker
We want to first play a game. We will show you a picture of an item. You need to give clues to your group, and get them to guess the item. You can't say the name of the item.
For example, if I have the word "water", I could say "thirsty".
[*Ask colleague to show first item to respondent 1*].
[*Repeat for all 3 respondents.*]

2b. General discussion
1. If you could choose any film star to deliver groceries to your home, who would you choose? *Why? Do you agree/disagree?*
2. What do you think of the working conditions for Swiggy & Zomato workers? *Why? Do you agree/disagree? Can you give me an example?*
3. How do you think companies should improve safety for delivery workers? *Why? Do you agree/disagree? Can you give me an example?*

3. Hiring discussion
- So far, different videos have been shown to different participants. This way, we can understand what people know about worker rights and consumer rights.
- We would like to hear your thoughts on our delivery options. We want you to discuss the **advantages** and **disadvantages** of each option.
- You should consider 3 things: (i) **Person** and their details (ii) **Items**, (iii) **Conversation** for 15-minutes.
- You will then make a **collective choice** to decide which option you prefer.
- In the first round, you will do this for 4 pairs. If one of these pairs is selected by the scratch-card, you will **all actually receive a delivery** from that person.
- Since you are all in the same location, this makes it **easier for us to organize** the deliveries.
- Later, you will each do **6 more choices individually** and in private.
  Do you have any questions?
- *[Ask assistant to give out choice sheets.]*

For each pair (x4)
- **Introduction questions:** (do not mention "transgender")
  - What are the differences between A and B?
  - What are the advantages/disadvantages of A/B?
  - What do you think about the photos?
  - What do you think about the details?
  - What do you think about the items?
- **Prompts** (*ask questions that cannot be answered with yes/no, do not mention "transgender"*)
  1. *"Why do you think that?"*
  2. *"What's your take on this?"*
  3. *"Can you talk about that more?"*
  4. *"Help me understand what you mean"*
  5. *"Can you give an example?"*
  6. *"What else do you think about this?"*
  7. *Ask them to ask each other what they think - "Don't tell me, tell it to your neighbours"*

- **[2-person discussion only, to the listener]**
  - Did you hear what other people chose?
  - Did you ask what the reason was?
  - [If they say no – ask speakers to repeat their preferences and reasons to the listener.]

I Video scripts

Rights video scripts

Slide 1:

Hello. In this video, we are going to talk about what rights people have when you purchase groceries from someone.

Slide 2:

Translation of slide text:
What are rights?

First, let's talk about what rights are. Rights are rules about what people are entitled to.

Slide 3:

Translation of slide text:
Consumer rights

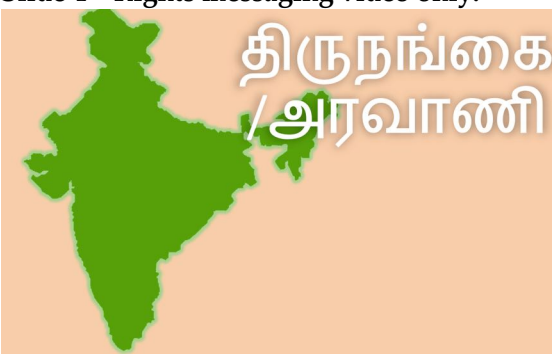
For example, when you make a purchase, you have the right to complain if the purchase is unfair. This is called your consumer right.

Slide 4 – Legal rights video only:

Translation of slide text:
Transgender people

As another example, the Supreme Court of India, the most powerful legal institution in the country, gave the thirunangai people all the fundamental rights under the Constitution of India as others in 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.

Slide 4 – Rights messaging video only:



Translation of slide text:
Transgender people

As another example, 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.

Slide 4 – Control video only:

Translation of slide text:
Voting rights

As another example, 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.

Slide 5:

Finally, the delivery workers, such as those that work in Zomato or Swiggy, should have worker rights. For example, some delivery workers think that they should have the right to employment benefits such as ESI/PF/Insurance.

J Data and measurement

J.1 Predicted choices (community)

Participants first made incentivized predictions of the choices of others in the study whom they did not know. They were shown 3 pairs of delivery options, and truthfully told that 20 other people in the study in the participants' area had been shown those same pairs. They had to predict how many of those 20 picked each option.

If they made the closest guess on average across all 3 pairs, they were entered into a lottery to win 3000 Rs.' worth of additional items. 2 of the 3 pairs were male-to-male comparisons, and 1 pair compared a male and a transgender.

A randomly selected half of the participants were always asked how many picked the option on the left, and half were asked how 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.

J.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. If they correctly guessed all 4 combinations (2 predictions for 2 group members each) 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.

J.3 Social desirability score

To measure the social desirability score of each participant in phase 1, I use an adapted version of the Crowne & Marlowe (1960) 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 the basic specification in Table A46.

First, I correct the score for acquiescence bias, or the tendency to agree with whatever question is being asked. This correction is common in the psychometric literature and has been shown to substantially improve the reliability of psychometric constructs, including in developing country contexts (Soto et al., 2008; Rammstedt & Farmer, 2013; Laajaj & Macours, 2019). To make this correction, I take the following steps:

1. Reverse the reverse-coded items.
2. Take the average of all positively-coded items for each individual i.
3. Subtract this from the average of the reverse-coded items for the same individual i.
4. Divide this by two to get the acquiescence score AS_i for individual i.
5. Correct individual i's raw scores by adding AS_i to every reverse-coded item, and subtracting AS_i from every positively-coded item.

Second, I calculate a social desirability score based on weights from a factor analysis that assumes a single factor. The loadings for each of the 7 variables are: (0.23, 0.03, 0.36, -0.13, 0.35, 0.31, 0.32) . Following the psychometric literature (e.g., Rammstedt & Farmer, 2013), I remove measures with a loading less than 0.3, and weight the remaining measures with the factor loading.

Third, I use construct an index based on inverse covariance weights, as seen in (Anderson, 2008).

J.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 John & Orkin (2022). 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.

That 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.

J.5 List experiment

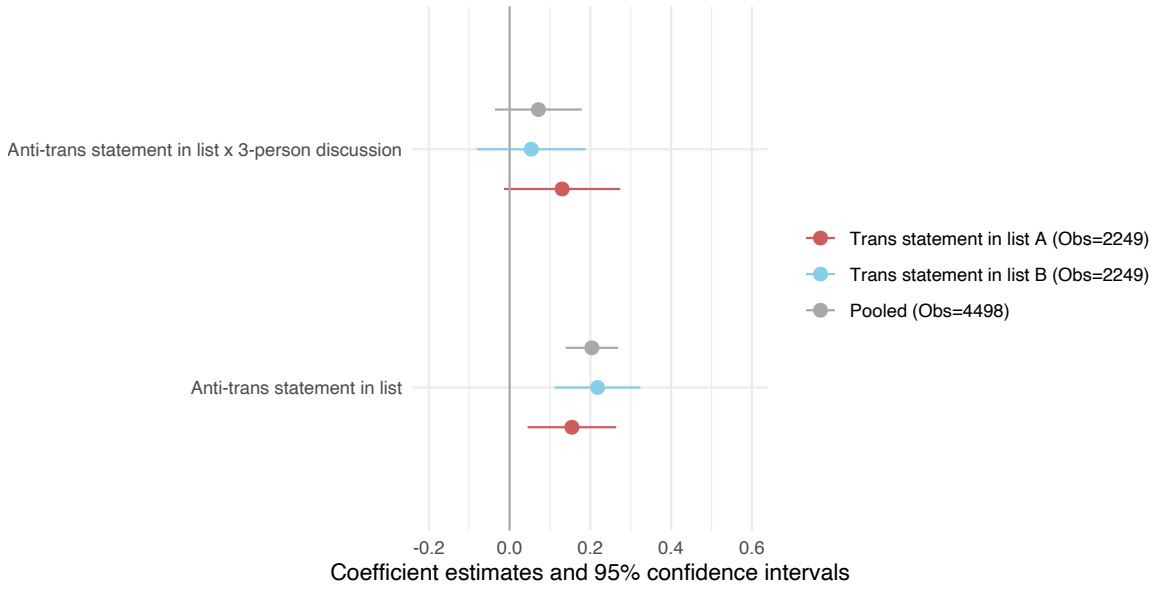
To measure negative attitudes towards transgender people, I use a double list experiment (Droitcour et al., 2004; Glynn, 2013). In this method, participants are shown two lists of statements (list A and list B), and are asked how many statements from each list they agree with. They are not asked *which* 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, shown in Appendix Table J1. For each participant, either list A or list B is randomly selected to include one additional statement: "In general, if I see a transgender, 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 (Chuang et al., 2021). 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, Appendix Figure J2 shows that the treatment effect estimates are quantitatively very similar (0.130 and 0.054 respectively. The difference between the estimates is not significant (p=0.93).

Table J1: *List experiment statements*

List	#	Question
List A	1	In our household, we often buy deliveries of goods online
	2	I can speak English
	3	I prefer rice to dal
	4	I am non-vegetarian
	5	It is easy for me to order things using an app
List B	1	I prefer to watch news on my mobile than on TV
	2	I think men generally talk more than women
	3	I would never buy more than 500 Rs. of groceries in one go
	4	The quality of vegetables is the most important factor when buying vegetables
	5	I prefer coffee to tea

Figure J2: *Treatment effect on list experiment does not depend on which list is used to estimate it*



Notes:

J.6 Group relations

We asked participants questions about their relationships to 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:

1. What is your relation with [*NAME*]?
2. 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:

3. How long have you known [*NAME*]? (Options: *Less than 6 months, 6 months to 1 year, 1-5 years, 5+ years*)
4. 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*)
5. How often do you ask [*NAME*] for advice? (Same options as 4)
6. How often do you ask [*NAME*] for recommendations of items to buy? (Same options as 4)
7. How frequently do you tell secrets to [*NAME*]? (Same options as 4)

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. The estimated loadings are in Table J3. 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.

Table J3: *Loadings for group relations index*

ID	Question	Loading
1a	Neighbour (=1)	-0.40
1b	Friend (=1)	0.19
1c	Close family (=1)	0.23
1d	Other family (=1)	0.22
2	How well do they know?	0.66
3	How long have they known?	0.49
4	Frequency: talking	0.59
5	Frequency: asking advice	0.73
6	Frequency: asking recommendations	0.72
7	Frequency: tell secrets	0.61

J.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.[45] 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

[45] 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.

anonymous.[46] 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.

J.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:

1. Out of 10, how confident is [*NAME*]?
2. Out of 10, how quiet is [*NAME*]? (*reverse coded*)
3. Out of 10, how like a leader is [*NAME*]?
4. Out of 10, how shy is [*NAME*]? (*reverse coded*)
5. Out of 10, how talkative is [*NAME*]?
6. Out of 10, how admirable is [*NAME*]?
7. 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 with the following steps.

1. I correct for acquiescence bias in the same way as described in Appendix J.3.
2. I use a factor analysis with one factor to generate loadings for each of the 7 measures. The estimated loadings are (0.47, 0.6, 0.36, 0.7, 0.61, 0.36, 0.44) . Since all loadings are above 0.3, I retain all the measures and create an index using a weighted sum of all measures, where the weights are proportional 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 0.23, $p<0.001$), 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.

J.9 LASSO controls

Following Belloni et al. (2014), I use double LASSO to select controls in the main results. The full set of possible controls that were selected from are in Table J4. In addition:

[46] 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.

- In interaction specifications where the main treatment was identified by the interaction *Worker is trans* × *Treatment*, I also include the controls interacted with *Worker is trans* as possible controls.
- I 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 Table 1 and 2 by the LASSO selection process in Table J5.

Table J4: *All potential controls used in LASSO control selection process*

Variable
Female (=1)
Speaks English (=1)
Reads English (=1)
Hindu (=1)
Bachelor's degree (=1)
Married (=1)
Employed (=1)
Landlord (=1)
Num. children
Employer (=1)
Household size
Monthly household food expenditure per capita (Rs.)
Num. family members in group-of-3
Num. neighbours in group-of-3
Num. friends in group-of-3
Taken part in market research survey (=1)
Has received free item as promotion (=1)
Someone in household ordered taxi with app (=1)
Someone in household ordered food with app (=1)
Someone in household ordered other items with app (=1)
Self-reported WTP for delivery
Respondent would normally be household member who receives delivery (=1)
Relative number of items offered by worker
Relative reliability score
Reliability score is shown (=1)
Reliability score of the benchmark worker

Table J5: *LASSO controls used in Table 1 and Table 2*

	Effect of 3-person discussion (Table 1)		Phase 2 effects (Table 2)	
Variable	(2)	(3)	(2)	(3)
Female (=1)	X		X	X
Group-level control: Bachelor's degree (=1)			X	
Group-level control: Employed (=1)			X	
Group-level control: Employer (=1)	X	X	X	
Group-level control: Has received free item as promotion (=1)			X	X
Group-level control: Household size			X	
Group-level control: Landlord (=1)	X			
Group-level control: Married (=1)			X	
Group-level control: Reads English (=1)			X	
Group-level control: Relative number of items offered by worker	X		X	
Group-level control: Relative reliability score			X	X
Group-level control: Reliability score of the benchmark worker			X	
Group-level control: Self-reported WTP for delivery	X		X	
Group-level control: Someone in household ordered food with app (=1)	X			
Group-level control: Taken part in market research survey (=1)	X		X	X
Has received free item as promotion (=1)			X	
Married (=1)	X		X	
Reads English (=1)			X	X
Relative number of items offered by worker	X	X	X	X
Relative reliability score	X		X	
Self-reported WTP for delivery			X	X
Someone in household ordered other items with app (=1)			X	
Taken part in market research survey (=1)			X	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 household 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.