Pre-Analysis Plan:
Speaking Up For Their Rights: Reducing Anti-Transgender Discrimination in India

Duncan Webb, Paris School of Economics

March 6, 2023

1 Intro

This project seeks to measure how anti-transgender discrimination is affected by information about transgender rights and group discussions about hiring transgenders.

2 Sample

The total sample size will be approximately 2250. Participants will be recruited from up to 6 districts in Tamil Nadu. Participants in all treatment arms will be recruited in groups of 3. All members of a group will live in the same community, and be of the same gender. Same-gender groups make recruitment easier, and reduce the probability of a single (male) group member dominating a group discussion. The members of a group will carry out the survey simultaneously. Subjects will be recruited by enumerators using a number of strategies, including direct household canvassing and introductions from community leaders such as village panchayat heads and anganwadi center workers. Enumerators will be blind to treatment status while they are recruiting participants. The sample will include both men and women from both rural and urban areas, and will be limited to individuals who can read Tamil and are between the ages of 20 and 65.

3 Experimental design

3.1 Methodology overview

The main study takes place over the course of one session with the participant (the "main survey"). The purpose of the survey is to measure the effect of infor-

mation about transgender rights and participation in group discussions on the level of hiring discrimination against transgenders. To measure discrimination, participants will be shown a series of options for grocery deliveries that include a choice of worker. Their choice will determine which delivery option they receive for free in approximately 1 week's time. Approximately 1 week later, a member of the survey team will carry out this delivery and elicit a follow-up survey.

3.2 Treatments

There are two main treatment variations.

First, at the start of the main survey, participants are shown one of three different videos:

1. **Control video** - gives no information about transgender rights, and includes placebo information about voting rights in India
2. **Messaging video** - the participant is shown persuasive messaging, arguing that transgenders in India should have fundamental rights
3. **Law video** - the participant is shown a video that tells them that transgender people actually do have *legal* rights in India because of a Supreme Court ruling in 2014

All information is embedded in a longer video that discusses consumer and worker rights in the context of delivery services. This helps obfuscate the purpose of the video.

Then, participants take part in the first round of hiring ("Round 1"). In this round, they carry out a series of 4 incentivised choices over which delivery worker they would like to hire.

In round 1, participants are cross-randomised into one of two conditions:

1. **Discussion** - respondents take part in a discussion among their group of 3, in which they discuss which workers they prefer and why, and then make joint choices.
2. **No discussion** - round 1 is answered individually and in private.

In the main treatment arms, all participants in a group see the same video, yielding the following 6 arms, with approximately 300 individuals in each arm.

1. Control video + No Discussion
2. Messaging video + No Discussion
3. Law video + No Discussion
4. Control video + Discussion
5. Messaging video + Discussion
6. Law video + Discussion

In an additional 7th treatment arm, containing approximately 450 individuals, one person per group is shown the *Law* video, while the other two people are shown the *Control* video. All groups in this arm are involved in a discussion:

7. Mixed videos + Discussion

This arm provides independent variation in a participant's treatment status, conditional on the treatment status of others in their group, permitting the identification of within-group spillover effects.

I use $k \in \{1, ..., 7\}$ to denote each of the 7 treatment arms.

After round 1 of hiring, participants carry out round 2 of hiring, in which they

make 6 more choices of which delivery worker they would like to hire. Choices in round 2 are made individually and in private.

In total, across rounds 1 and 2, participants make 10 binary choices of delivery worker. Scratch-cards will be used to randomly select 1 of the 10 hiring choices to be implemented, so that the participant actually receives a delivery from the chosen worker 1 week later. To minimise risk to transgender workers, the randomisation will be designed so that choice pairs that include a transgender worker will be selected in less than 1% of cases. In the other cases, the randomisation will select pairs that include members of the team of enumerators as workers.

"High-stakes" condition. Approximately 50% of the sample in treatments 1-6 will be randomised into a "high-stakes" condition. In this condition, participants will be told that they will not only receive 1 delivery from their chosen worker, but instead will receive 3 deliveries from the same worker over the course of the following 3 months. This variation will be used to examine whether reducing the probability that participants engage in "cheap talk" or socially desirable behaviour affects the treatment effect size. For all main analyses that do not focus on this variation, the high-stakes and low-stakes conditions will be pooled. I will separately analyse heterogeneity along these lines as a supplementary analysis.

Stratification. Randomization will be stratified by (i) participant gender, because because piloting demonstrated that gender was an important source of heterogeneity, and by (ii) the survey team that carries out the first survey. The survey team is used to stratify because it affects the hiring choices a participant sees. When a participant is making hiring choices, the pool of delivery workers she sees is composed of survey enumerators who will subsequently carry out

the follow-up survey and delivery. To ensure impartiality, the enumerator conducting the initial survey will be excluded from the pool of potential delivery workers for that survey. Since the enumerator team that carries out the first survey affects the set of worker profiles the participant sees, the team is a source of heterogeneity and will therefore be used to stratify.

Randomization will be stratified by team using IDs that are specific to a (team $\times$ participant-gender) combination. IDs are pre-randomized into each condition. Teams will descend through the list of IDs in the order in which they carry out surveys. The list will be divided into blocks of 7 IDs (corresponding to the 7 treatment groups). To account for the increased sample size in the 7th treatment group, one in every two blocks will include an additional 8th ID that is assigned to the 7th treatment group. The order of treatments within each block is randomly assigned. As long as each team completes a whole number of blocks of IDs, the treatments will be balanced within a given team-gender combination. Even if they do not complete a whole number of blocks, the treatments will be approximately balanced within a team.

4 Outcomes

4.1 Primary outcome

Discrimination. After round 1, participants will take part in round 2 of hiring. In this round, they will be faced with 6 binary choice-pairs, in which they choose a combination of delivery worker and items. The number of items offered by each worker will be randomly varied; sometimes one worker in a pair will offer more items than the other. In each pair, the "Benchmark" choice is always male, while the "Comparator" choice can be male, female,

or transgender. Anti-trans discrimination is measured as the reduction in the probability that the Comparator is chosen when the Comparator is transgender.

The main specification for person i and worker pair j is:

$$\begin{aligned} ChooseComparator_{ij} = &\sum_{k=2}^{7} \beta_k (Treat_{ik} \times Trans_{ij}) + \gamma Trans_{ij} + \sum_{k=2}^{7} \delta_k Treat_{ik} \\ &+ \mathbf{X}'_{ij}\Gamma_0 + (\mathbf{X}'_{ij}\Gamma_1 \times Trans_{ij}) + \lambda ComparatorLeft_{ij} \\ &+ \pi_{s(i)} + (\pi_{s(i)} \times Trans_{ij}) + r_j + \varepsilon_{ij} \end{aligned} \tag{1}$$

where:

- $Trans_{ij} = 1$ if the pair j shown to i includes a transgender individual, and is 0 if the pair j only includes cis-men or cis-women
- $Treat_{ik} = 1$ if individual is in treatment group k, and is 0 otherwise, where $k \in \{1, ..., 7\}$ corresponds to each of the 7 treatment arms described above.
- $\mathbf{X}_{ij}$ is a vector of controls, chosen following the strategy in Section 5.1. In this specification they are interacted with $Trans_{ij}$ in order to control for differences in discrimination driven by observables.
- $ComparatorLeft_{ij}$ takes the value 1 when the comparator appears on the left of the screen in pair j shown to i, and 0 if it appears on the right of the screen
- $\pi_{s(i)}$ are stratum fixed effects, which are interacted with $Trans_{ij}$ in this specification in order to control for stratum level differences in discrimination across all treatment groups
- r_j denote pair-order fixed effects

The main treatment effects are measured by the coefficients β_k, which describe the changes in discrimination caused by the treatments. γ describes the base-

line level of discrimination against transgenders in the hiring choices in the $(\textit{Control video}, \textit{No discussion})$ group.

In the main specification, I will only use the choices from round 2, and not use any choices from round 1 in the analysis. For *No Discussion* participants, I will test whether there is a significant difference in treatment effects between round 1 and round 2 decisions. If there is no detectable difference, I will pool round 1 and round 2 decisions in order to increase statistical power on the estimates of β_k.

Since there is only one primary outcome, I will not adjust for multiple hypothesis testing for the discrimination measure.

4.2 Secondary outcomes

Discrimination:

1. **Follow-up discrimination**. In the follow-up survey, respondents will be asked to choose delivery options for 6 choice-pairs. Unlike in the main survey, this choice will be explicitly hypothetical. I will use an analogous specification (based on equation 1) to analyse whether changes in discrimination persist after 1 week. For this analysis, I will drop the small subset of individuals who actually receive a delivery from a transgender individual.

Attitudes:

1. **Implicit Association Test (IAT)**. Enumerators will administer an IAT that measures the implicit association of male and transgender faces with positive and negative words. I will calculate the D-score for each individual based on the algorithm seen in Greenwald et al. (2003). The

D-score is a standardised proxy for implicit bias, based on calculating the difference in response times between pro-stereotypical modules and counter-stereotypical modules, normalised by the standard deviations of those modules. I will analyse treatment effects using an OLS regression of the D-score on treatment dummies (a dummy for each of the 7 treatments), controlling for whether the participant saw the counter- or pro-stereotypical module first (this order is randomised).

I will test for the validity of the IAT responses across the whole sample and *only* include the IAT as a secondary outcome if the proportion of responses which pass a number of validation checks is sufficiently high. These validation checks will include:

- Proportion of respondents who enumerators rated as having difficulty responding or not understanding the instructions
- Proportion of responses with overly long response times
- Proportion of responses with overly short response times
- Proportion of IAT modules having to be restarted or skipped due to technical errors

2. **List experiment**. I will measure the proportion of people who agree with the statement "I would not want to have a conversation with a transgender person" using a double list experiment. Enumerators will read out two lists (A and B) that contain 6 non-sensitive statements about a participant's preferences. The statement about transgenders will be randomly added to either list A or list B. Whether list A or list B is read first will also be randomised. For each list, enumerators ask how many statements in the list they agree with. I will analyse treatment effects using an OLS

regression of the number of statements participants agree with on the interaction between treatment-dummies (one for each of the 7 treatments) and a dummy indicating whether the trans statement is included in the list, controlling for list and list-order fixed effects.

I will test for the validity of the list experiment responses across the whole sample and only include the list experiment as a secondary outcome if the pattern of responses passes a validation check that makes use of the *double* list experiment set-up. Because there are two lists, I can calculate two separate estimates of the proportion of people who agree with the statement about transgenders: one for list A, and one for list B (Chuang et al., 2021). The results will be deemed invalid and not used as a secondary outcome if there are significant differences between the estimates from list A and list B.

3. **Disapproval of discrimination**. I will measure whether participants disapprove of apparent discrimination against transgender individuals. Respondents will be shown a hypothetical scenario that strongly suggests discrimination. They will be shown a pair of delivery workers, one of whom is a clearly high-quality transgender worker (with a training score of 9, offering 2 items), and the other as a clearly low-quality male worker (with a training score of 5, offering 1 item). Respondents will be asked to imagine that someone in their locality chose a low-quality male partner, and then asked if, in the respondents' opinion, that person did something wrong. I will analyse the effects of the treatments using an OLS regression, using an indicator of whether the respondent said the choice was wrong as an outcome variable, using treatment dummies (one for each of the 7 treatments) as explanatory variables, and controlling for order fixed

effects, and fixed effects for whether the low-quality person is on the left or right. As a placebo test, I will also examine whether there are treatment effects on an equivalent question that compares a high-quality woman to a low-quality man.

Norms:

1. **Predicted choices (community)**. Participants will make incentivised predictions of others' hiring choices. For 3 delivery option pairs, they will be told that 20 other people in their locality were shown those delivery-option pairs, and they have to estimate how many of those 20 picked one of the options (the option asked in the question is randomly selected from the pair). If they make the closest guess on average across all 3 pairs, they will have a chance of winning a prize hamper. 2 of the 3 pairs will be male-to-male comparisons; the third pair will compare a male and a transgender. For analysis, the outcome variable will be the estimated proportion of the 20 individuals who selected the option on the side that the transgender worker appears, and the explanatory variable will be an interaction of the treatment dummies (one for each of the 7 treatments) with a dummy indicating that the pair includes a transgender worker. I will control for whether the transgender worker appears on the left or right.

2. **Predicted choices (own group)**. Participants will also make incentivised predictions of the hiring choices of the two other people in their group. For each of the two other group members, they will be asked to predict which option they chose for two choice-pairs. One choice-pair per person will compare a male and a transgender. If they correctly guess all 4 combinations they will be entered into a lottery to win a separate prize. For

analysis, the outcome is whether they predict that their group member will choose the *Comparator* (as defined above), and the explanatory variables are treatment dummies (one for each of the 7 treatments) interacted with whether the *Comparator* is transgender.

Beliefs:

1. **Perceived similarity**. To measure whether participants perceive transgenders to be more similar to them after being exposed to discussions and videos about rights, I will construct a *perceived similarity index*.

   First, respondents will be asked to report their preferences regarding (i) who they would call in an emergency, (ii) how much importance they place on their family's preferences, (iii) how much they value education, and (iv) where they would be happy to go for social gatherings. Later in the survey, they will be shown a photo of a transgender individual and be asked to predict that individual's responses for the same question. They will also be asked two further questions about the transgender person: (v) whether the person has similar preferences to them, and (vi) whether the person is likely to use a delivery service. Question (vi) will not be used in the analysis, and is used to obfuscate the purpose of the study.

   The index will be constructed by:

   (a) Constructing binary indicators of whether respondents predicted that the transgender individual gave the same answers as them for questions (i) to (iv), and a binary indicator for (v) of whether they said that the person has similar preferences.

   (b) Combining these 5 indicators into a summary index, following Anderson (2008).

The index will be used as an outcome variable in a regression with treatment dummies as the explanatory variables (one dummy for each of the 7 treatments).

2. **Beliefs about the law**. I will analyse whether three measures of people's beliefs about the law regarding transgenders are affected by the information videos:

   (a) *Knowledge of law*: whether they say that they know about a law in India that relates to transgenders *and* correctly describe the rights they were given

   (b) *Discrimination is breaking the law*: whether they say that someone that hires a man over a better-qualified transgender individual is breaking the law

   (c) *Discrimination can lead to legal consequences*: whether they say that someone that hires a man over a better-qualified transgender individual could be sued, fined, or imprisoned

   These will be combined into an index of beliefs about the law. This index will be regressed on treatment dummies. For the main analysis on this beliefs-about-the-law index, I will pool *Discussion* and *No Discussion* arms and analyse the pooled treatment effects of the *Messaging video* and the *Law video* compared to the *Control video*. For a secondary analysis, I will examine the effects of all treatment groups 1-6 separately.

Discussion dynamics (only for treatment groups 4-7, i.e., those with *Discussion*)

1. **Discussion discrimination**. In round 1, participants will make collective choices for who to hire in their groups of 3. As a primary analysis, I will analyse the treatment effects on these group choices using a specification

analogous to equation 1. As a secondary analysis, I will also compare the round 1 choices in the discussion conditions to those in the non-discussion conditions by regressing round 1 choices on the interaction between $Trans_{ij}$ and the full set of treatment dummies 1-7.

2. **Probability of endorsing**. Group discussions will be recorded and transcribed. Only recordings with sufficiently high data quality will be kept for the analysis; recordings with a high proportion of inaudible speech will be dropped. An enumerator will record who speaks at each time during the discussion, so that each person's statements can be attributed to them. Research assistants will code a binary variable in the transcript, $Endorse_{aijt}$, that indicates that the participant i made a statement at time t in the discussion advocating that the group choose candidate a when deciding over a pair j. This will include both unjustified endorsements of candidate a (e.g., "I think we should pick A"), and positive justifications for choosing a (e.g., "A appears to be better qualified").

   The main reduced-form analysis will test whether the treatments impact the probability of a person endorsing a transgender candidate at least once during the discussion. In secondary analyses, I will examine whether the treatments make a participant more likely to endorse a transgender candidate *early* in the discussion. Does the participant endorse a transgender candidate before she has endorsed any other candidate? And does she endorse a transgender candidate before others make a statement in the discussion?

3. **Spillover analysis**. For this analysis, only the control video group, the mixed video group and the law video group (all with discussions) will be included. Let $Law_i = 1$ if i saw the law video, and $N_{-i} := \sum_{q \neq i} Law_q$ be

the number of other people in i's group who saw the law video.

To analyse within-group spillovers, I will use two basic specifications.

First, assuming linear spillover effects, I will estimate:

$$\begin{aligned} ChooseComparator_{ij} = \beta(Law_i \times Trans_{ij}) + \gamma(N_{-i} \times Trans_{ij}) + \\ \delta_1 Trans_{ij} + \delta_2 Law_i + \delta_3 N_{-i} + \mathbf{X}'_{ij}\Gamma_0 + (\mathbf{X}'_{ij}\Gamma_1 \times Trans_{ij}) + \\ \pi_{s(i)} + (\pi_{s(i)} \times Trans_{ij}) + \varepsilon_{ij} \end{aligned} \tag{2}$$

(using the individual choices in round 2 as the outcome $ChooseComparator_{ij}$).

As a secondary analysis, allowing for more flexibility, I will estimate a saturated model using dummy indicators for the following 4 treatment groups: (i) individuals in the control video group (the omitted category), (ii) individuals in the mixed video group who saw the control video, (iii) individuals in the mixed video group who saw the treatment video, and (iv) individuals in the treatment group. These dummy indicators will be interacted with $Trans_{ij}$ as in the previous specifications.

In a supplementary analysis, I will also measure the causal effects of having someone else in a group endorse a transgender in the discussion by instrumenting this endorsement with other group members' treatment status (Law_i). I will measure the effect of being exposed to endorsement from others on a respondent's own propensity to endorse (a "cascade effect"), and on the respondent's level of discrimination.

4. **Enumerator observations**. I will analyse changes in the discussion dynamics by measuring whether there are treatment effects on enumerator observations of the discussion dynamics, and the differential effects on discussions of choice-pairs that include transgenders. The outcomes will

include: whether the word "transgender" was mentioned, whether someone said something positive or negative about transgenders, how much discussion occured (both for a given choice-pair and overall).

I will use a model that is in development at the time of writing to carry out further empirical analyses on the group discussion dynamics.

5 Empirical specification

5.1 Controls

For all regressions, I will use double LASSO (Belloni et al., 2014) to select the controls included in the specification. The list of potential controls include:

- *Individual covariates*, measured before exposure to information treatments: religion, education, marital status, employment status, landlord status, whether participant has a child in the household, number of children in household, number of children below the age of 10 in the household, whether participant has employed someone, smartphone ownership, household size, per capita food expenditure, Crowne and Marlowe (1960) social desirability score, a measure of willingness to persuade in discussions, a proxy for baseline progressive social attitudes, whether they have used an app or website to order a taxi/motorbike/auto, whether they have used an app or website to order food/drink/groceries, whether they have used an app or website to order other goods, whether they normally receive groceries themselves, willingness to pay for delivery (i) in 7 day's time, and (ii) today.
- *Group-level covariates*: For each of the individual covariates, I will use the mean for the other two members of the participant group as a possible

control. In addition, possible controls will be added for the relationship between the group members.

- *Worker profile covariates*: the difference between the number of items offered by the *Comparator* and the number of items offered by the *Benchmark*; the worker-training score shown for the *Benchmark* worker; the difference between the worker-training score for the *Comparator* worker and the *Benchmark* worker; the number of quality signals included in the worker profile (where applicable); the values of the additional quality signals included in the worker profile (where applicable); whether the *Comparator* is a woman.

For specifications based on interaction effects, such as the specification in equation 1, I will also include the interaction between $Trans_{ij}$ and the above controls in the list of potential controls.

Experimental design covariates: Where applicable, specifications will include fixed effects to account for features of the experimental design, including order fixed effects. These are specified in Section 4.2 where applicable. These will be included in the specification, rather than being selected by the LASSO process.

5.2 Inference

Standard errors on regressions will be clustered at the group-of-3 level. I will perform randomization inference as a robustness check.

Multiple hypothesis testing. I will correct for multiple hypothesis testing by additionally reporting the q-values adjusted for the false discovery rate, following the procedure in Anderson (2008). I will apply the correction within the following sets of secondary outcomes on the main specification reported:

- *Attitudes*: {IAT, list experiment, disapproval of discrimination}
- *Norms*: {Predicted choices (community), predicted choices (own group)}

As noted above, I will not test for multiple hypothesis testing for the discrimination outcome since it is the only primary outcome. If the IAT or list experiment outcomes fail the validation checks described above and are therefore excluded from the main analysis, they will also be excluded from the multiple hypothesis testing correction.

6 Heterogeneity analyses

I will run a number of heterogeneity analyses along the following dimensions in order to understand the mechanisms behind my results:

1. *High-stakes condition*, i.e. whether participant was randomised into the high-stakes condition with additional deliveries from the same worker. For this comparison, I will pool *Discussion* and *No Discussion* arms, leaving 6 comparison groups: $\{\text{Control video}, \text{Messaging video}, \text{Law video}\} \times \{\text{High-stakes}, \text{Low-stakes}\}$.
2. *Worker quality signals*, i.e. whether anti-transgender discrimination is reduced when the worker profile includes more details indicating the quality of the workers (namely, their score in a training task, whether they speak English, the number of years of experience), and the values of those details
3. *Number of items offered*. Does the level of anti-trans discrimination change when the transgender option is offering 1 or 2 additional grocery items, rather than the same number of grocery items?

4. *Social desirability bias score*, as measured by an index elicited before treatment based on the Crowne and Marlowe (1960) social desirability index.

5. *Baseline progressive attitudes*, as proxied by an index elicited before treatment that asks about progressive social attitudes

6. *Perceived purpose of the experiment*, i.e., whether participants correctly guess that the study's purpose is to measure anti-transgender discrimination.

7. *Demographic characteristics*, including:

    - Participant gender, which may be correlated with baseline attitudes and the dynamics of group discussions
    - Religion, since transgenders have a specific cultural role in Hinduism that may exacerbate or mitigate discrimination
    - Whether the participant is a landlord, in order to examine whether discrimination is common among those who would be in a position to provide transgenders with housing
    - Whether the participant has children in the household, to explore whether those with children fear exposure to transgenders more
    - Socioeconomic status (SES), proxied by household per capita food expenditure and education level: discrimination may be less severe among low-SES people if the cost of sacrificing more grocery items in order to discriminate is greater, but it may be more severe if low SES is correlated with less progressive attitudes

8. *Group composition*. I will measure whether there are stronger effects in the mixed-video group if the treated individual in the group is a persuasive "thought-leader", as proxied by an above median score on an index of

willingness to persuade others.

7 Supplementary analyses

I will run a number of supplementary analyses.

Using the main sample, I will analyse:

1. *Salience*. Do the treatments increase the salience of transgenders as measured by a participants' ability to recall the word transgender from a list of words (conditioning on their overall recall level for the list of words)?

2. *Effect on choosing women*. I will analyse whether the information treatments and group discussion have an effect on a participant's propensity to choose a woman over a man, in addition to changing the propensity to choose a transgender over a man. A positive spillover effect of this kind would indicate that discussions and information about fundamental rights can lead to *generalised* reductions in discrimination, even when those discussions and information do not pertain to the group in question. Including women as worker options also helps obfuscate the purpose of the study and reduces the risk of experimenter demand effects.

3. *Effect on perceived reliability*. Do the information treatments and group discussion have an effect on the perceived reliability of transgender individuals, as measured by the question: "If this person were chosen to deliver the items, how likely do you think this person would be to complete the delivery?"

Bibliography

Michael L. Anderson. 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, December 2008. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214508000000841.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650, April 2014. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdt044.

Erica Chuang, Pascaline Dupas, Elise Huillery, and Juliette Seban. Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, 148:102582, January 2021. ISSN 03043878. doi: 10.1016/j.jdeveco.2020.102582.

Douglas P Crowne and David Marlowe. Marlowe-Crowne social desirability scale. *Journal of Consulting Psychology*, 1960.

Anthony G. Greenwald, Brian A. Nosek, and Mahzarin R. Banaji. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85:197–216, 2003. ISSN 1939-1315. doi: 10.1037/0022-3514.85.2.197.